

Raamwerk Evaluatie Interventies Leven Lang Ontwikkelen

Citation for published version (APA):

de Grip, A., Künn, A., & Montizaan, R. (2019). *Raamwerk Evaluatie Interventies Leven Lang Ontwikkelen*. ROA. ROA Reports No. 002 <https://doi.org/10.26481/umarep.2019002>

Document status and date:

Published: 25/02/2019

DOI:

[10.26481/umarep.2019002](https://doi.org/10.26481/umarep.2019002)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Raamwerk Evaluatie Interventies Leven Lang Ontwikkelen

Andries de Grip
Annemarie Künn-Nelen
Raymond Montizaan

ROA Rapport

ROA-R-2019/2

Researchcentrum voor Onderwijs en Arbeidsmarkt | ROA
Research Centre for Education and the Labour Market | ROA

Colofon

© Researchcentrum voor Onderwijs en Arbeidsmarkt (ROA). Niets uit deze uitgave mag op enige manier worden verveelvoudigd zonder voorafgaande schriftelijke toestemming van de directeur van het ROA.

Researchcentrum voor Onderwijs en Arbeidsmarkt

Postbus 616

6200 MD Maastricht

T +31 43 3883647

F +31 43 3884914

secretary-roa-sbe@maastrichtuniversity.nl

www.roa.nl

School of Business and Economics

Maastricht University

Vormgeving

ROA secretariaat, Maastricht

ISBN: 978-90-5321-578-4

januari 2019

INHOUD

1	Inleiding	1
2	Evaluatie interventies LLO	5
2.1	Grootschalig gerandomiseerd experiment	5
2.2	Trapsgewijs gerandomiseerd veldexperiment	7
2.3	Difference - in - Difference methode	10
2.4	Regression discontinuity design	13
2.5	Voor- en nameting	17
2.6	Rangschikking interventiemethoden	20
3	Focus op individuele leerrekeningen en leerambassadeurs	23
3.1	Individuele leerrekeningen	23
3.2	Leerambassadeurs	27
4	Monitoring trend LLO in Nederland	31
4.1	Huishoudpanels en werknemersvragenlijsten	31
4.2	Werkgeversvragenlijsten	35
5	Beoordeling evaluatiemethode van toekomstige projecten	37
5.1	Beslisboom en voorwaarden waaraan voldoen moet worden als voor een bepaalde evaluatiemethode wordt gekozen	37
5.2	Checklist voor goedkeuring projecten	41
	Literatuur	43

1 INLEIDING

Het belang van een leven lang ontwikkelen is goed doorgedrongen tot brancheorganisaties, werkgevers en grote groepen werknemers en staat ook hoog op de agenda van de overheid. Het blijft echter de vraag hoe ervoor gezorgd kan worden dat iedereen zich een leven lang blijft ontwikkelen én zich de juiste competenties eigen maakt. Dat hier veel over nagedacht wordt, is onder andere te zien aan de talrijke initiatieven rondom het 'Leven Lang Ontwikkelen' (LLO). Het betreft echter meestal pilots die niet of nauwelijks geëvalueerd (kunnen) worden, waardoor vaak het wiel opnieuw moet worden uitvonden en er geen goed beeld is van wat effectieve manieren zijn om het LLO te stimuleren. Het is nog onvoldoende vanzelfsprekend dat een initiatief gericht op het bevorderen van een leven lang ontwikkelen zodanig opgezet wordt dat een goede effectmeting mogelijk is. Onder meer door selectie van specifieke bedrijven of werknemers in dergelijke initiatieven zijn de effecten niet eenduidig te analyseren en worden de effecten ervan vaak ook overschat, omdat de scholing ook zonder de stimulans (interventie) wel had plaatsgevonden (het zgn. *dead-weight loss effect*).

Dat niet iedereen bereid is om zich te scholen blijkt onder andere uit de recente studie van Künn-Nelen et al. (2018).¹ Zo is de motivatie om aan training deel te nemen onder laagopgeleiden en 55-plussers relatief laag. Echter, ook voor deze groepen werkenden blijkt dat zij meer bereid zijn om een cursus of training te volgen als zij geen eigen financiële bijdrage hoeven te leveren en de training niet in eigen tijd hoeven te volgen.² Een individuele leerrekening zou deze hindernissen voor deze groepen kunnen wegnemen.

Künn-Nelen et al. (2018) laten ook zien dat lager opgeleiden veel leren van de feedback van hun collega's. Dit suggereert dat zij mogelijk meer open staan voor tips van collega's die op hetzelfde niveau werkzaam zijn, dan van de feedback van hun leidinggevenden. Dit maakt het interessant om te onderzoeken of leerambassadeurs³ de bereidheid van met name lager opgeleiden om aan een cursus deel te nemen kunnen verhogen.

¹ Künn-Nelen et al. (2018), *Leren onder werkenden met een kwetsbare positie op de arbeidsmarkt*, ROA-R-2018/5, Maastricht.

² Bij alle werkenden, met uitzondering van werkenden met een tijdelijk dienstverband, neemt de bereidheid om te trainen af als er een eigen bijdrage in tijd of geld moet worden geleverd.

³ Zie I. Coenen (2018), *Vakbondswerk maken van scholing en ontwikkeling*, *Zeggenschap*, 2018/3, p. 40-42.

Naast de individuele leerrekening en leerambassadeurs zijn er nog veel meer LLO-initiatieven, waarvan het de moeite waard is om de effectiviteit ervan te onderzoeken. Daarbij kan men denken aan het verstrekken van opleidingsbudgetten (al dan niet met eigen regie van de medewerker), leeradviseurs voor bedrijven en de verbetering van de ondersteuningsstructuur. Een andere mogelijkheid is om de effectmeting te richten op een specifiek aspect van een interventie. Bijvoorbeeld door de interventiegroep een leerrekening te geven waarop de middelen naar bijvoorbeeld 3 jaar verjaren, terwijl de ingezette middelen op de leerrekening van de controlegroep onbeperkt geldig zijn. Een belangrijke vraag die al deze interventies oproepen is welk effect zij hebben op het LLO en voor welke groepen dit effect optreedt.

In dit rapport presenteren wij een raamwerk met vijf evaluatiemethodes om de effecten van initiatieven op het gebied van een 'Leven Lang Ontwikkelen' op een verantwoorde wijze te meten. We laten een hiërarchie zien die uiteenloopt van een randomised control trial als de sterkste methode om de effectiviteit van interventies te meten tot een voor- en nameting, die wel meet of na een interventie het LLO toeneemt, maar geen causaal verband kan aantonen. De hiërarchie die we presenteren is gebaseerd op een drietal dimensies. Allereerst wordt de Maryland schaal gebruikt. Deze schaal is ontworpen door Farrington e.a. (2002) en beoogt de kwaliteit van onderzoek te bepalen op een schaal met vijf gradaties. Deze schaal gaat echter niet in op de interne en externe validiteit van het onderzoek. Omdat beide van belang zijn bij de beoordeling van voorstellen om interventies te evalueren, hebben we de mate van interne en externe validatie toegevoegd als dimensies om de kwaliteit van de evaluatiemethodes te beoordelen.

1. Grootchalig gerandomiseerd experiment.
2. Trapsgewijs gerandomiseerd veldexperiment.
3. Difference in difference (DD)-methode.
4. Regression discontinuity (RD)-design.⁴
5. Voor- en nameting.

In Hoofdstuk 2 worden de vijf bovengenoemde evaluatiemethodes voor interventies op het gebied van een 'Leven Lang Ontwikkelen' besproken. Aan de hand van een voorbeeldstudie beschrijven we telkens allereerst de methode zelf. Daarna bespreken we achtereenvolgens de voorwaarden, de mogelijkheden en beperkingen en de toepasbaarheid van de evaluatiemethodiek voor LLO-interventies.

Los van de gebruikte methode is het belangrijk helder te krijgen welke data nodig zijn om LLO-interventies te evalueren. Longitudinale data zijn noodzakelijk, behalve bij een regression discontinuity design.⁵ Welke uitkomstmaten idealiter gebruikt moeten worden bij de evaluaties, hangt af van de interventie zelf. Is de interventie bedoeld

⁴ In Hoofdstuk 2.6 wordt een onderscheid gemaakt tussen een RD-design met en zonder voormeting. Een RD-design met voormeting heeft de voorkeur boven een DD-methode.

⁵ Zie voor meer informatie Hoofdstuk 2.

om bewustwording rondom het thema LLO te creëren of om de cursusdeelname te vergroten?⁶ Voor relatief harde uitkomsten zoals cursusdeelname of het besteden van vouchers kunnen grote datasets zoals de Enquête Beroepsbevolking of administratieve data van een bedrijf of gebruikt worden. Administratieve data hebben als voordeel dat selectieve respons op vragenlijsten omzeild kan worden. Wanneer dergelijke administratieve data niet voorhanden zijn bieden vragenlijsten een uitkomst. Daarbij kan men denken aan papieren of digitale vragenlijsten, maar ook aan vragenlijsten die worden uitgezet via een app. Voor bepaalde uitkomstmaatstaven is overigens altijd een vragenlijst vereist. Dit geldt met name voor gepercipieerde uitkomsten en voorkeuren. In veel gevallen is een combinatie van administratieve (bedrijfs)data en data verkregen via vragenlijsten gewenst om een compleet beeld van de effecten van LLO-interventies te krijgen.

Omdat zowel de geprefereerde methodiek als de uitkomstmaatstaven afhankelijk zijn van de interventie die men wil evalueren, komen we in Hoofdstuk 3 met expliciete voorbeelden voor de aanpak om twee LLO-interventies te evalueren. Het gaat hierbij om de evaluatie van de effecten van individuele leerrekeningen en leerambassadeurs.

Naast het evalueren van concrete LLO-interventies is het ook van belang om de ontwikkeling van het LLO in Nederland meer in het algemeen te monitoren. Enerzijds schetst dit de verschuivende achtergrond waarbinnen specifieke interventies plaatsvinden. Anderzijds geeft dit een beeld van de voortgang die de verschillende groepen werkenden, bedrijfssectoren en regio's in ons land maken op het terrein van het LLO. In Hoofdstuk 4 presenteren we manieren om de ontwikkeling van het LLO in Nederland op systematische wijze te monitoren. We gaan hierbij in op de mogelijkheden binnen bestaande datasets. Deze datasets lenen zich bovendien voor het opzetten van grootschalige gerandomiseerde veldexperimenten.⁷

Hoofdstuk 5 vormt het afsluitende hoofdstuk waarin een beslisboom wordt gepresenteerd voor het kiezen van de meest geschikte evaluatiemethode en een checklist om te beoordelen of projectvoorstellen al dan niet voldoen aan de eisen voor het goed uitvoeren van effectmetingen in het kader van LLO-interventies.

⁶ Een voorbeeld voor uitkomstmaten gerelateerd aan (het introduceren van) leerambassadeurs of individuele leerrekeningen zijn gegeven in Hoofdstuk 3.

⁷ Zie de voorbeeldstudie van Schwerdt e.a. (2012) die beschreven wordt in Hoofdstuk 2.1.

2 EVALUATIE INTERVENTIES LLO

In dit hoofdstuk bespreken we de in de inleiding genoemde vijf verschillende methoden om de effecten van LLO-interventies te meten. Dit doen we door telkens eerst een algemene beschrijving te geven van de methode en een voorbeeldstudie te bespreken. Vervolgens bespreken we voor elke methode de voorwaarden waaraan moet worden voldaan om te komen tot een causale effectmeting (een *intern valide* meting). Daarnaast bespreken we de mogelijkheden en beperkingen en de toepasbaarheid van elke methode. Het hoofdstuk sluit af met een tabel waarin we de verschillende methoden rangschikken op basis van zowel de Maryland schaal als de interne en externe validiteit.

2.1 Grootschalig gerandomiseerd experiment

De eerste evaluatiemethode betreft een grootschalig gerandomiseerd veldexperiment. Indien LLO-initiatieven als een grootschalig gerandomiseerd veldexperiment opgezet worden is er per definitie sprake van een effectmeting. Hoewel men in eerste instantie vaak niet denkt aan een dergelijke experimentele opzet bij LLO-initiatieven, laten Schwerdt e.a. (2012) zien dat dit wel degelijk mogelijk is.

Voorbeeldstudie

Schwerdt e.a. (2012) evalueren het effect van de introductie van een scholingsvoucher in Zwitserland. Op basis van de Zwitserse variant van de Enquête Beroepsbevolking (de zogenaamde Schweizerische Arbeitskräfteerhebung, SAKE) hebben zij een gerandomiseerde steekproef getrokken van 1.422 personen.⁸ Deze personen, behorende tot de interventiegroep, ontvingen begin januari 2016 een scholingsvoucher. Daar stond een controlegroep van 9.099 personen tegenover die deze scholingsvoucher niet hebben gekregen. Er waren drie verschillende scholingsvouchers van 200, 750 of 1500 Zwitserse Francs. Deze drie varianten zijn gerandomiseerd verdeeld onder de interventiegroep. Ontvangers van de voucher mochten zelf weten welke scholing zij wilden volgen.⁹ De vouchers konden door de interventiegroep in de eerste helft van 2016 ingezet worden voor scholingsdoeleinden. Omdat er sprake was van een gerandomiseerde steekproef

⁸ De gerandomiseerde selectie van de interventiegroep is apart gedaan voor de 40 regio's. De enige restrictie was de leeftijdsrestrictie van 20 tot 60 jaar.

⁹ Achteraf bleek dat slechts 10% van de vouchers ingezet is voor cursussen gerelateerd aan hobby's.

en omdat niemand op de hoogte was van het lopende experiment, resulteerde een vergelijking van scholingsdeelname onder de twee groepen in een causale effectmeting van de scholingsvoucher. Naast scholingsdeelname is ook geanalyseerd of het ontvangen van (*intention-to-treat effect*) en gebruik van de scholingsvoucher (*treatment effect*) resulteerde in het volgen van andere scholing, in een hoger salaris en in een hogere kans op werk 1 jaar later. Dit was mogelijk omdat de vragen op deze gebieden standaard opgenomen zijn in de Zwitserse variant van de Enquête Beroepsbevolking, de SAKE. Bovendien was het mogelijk om te evalueren in hoeverre de voucher geleid heeft tot *dead-weight loss*. Door het grootschalige karakter van het experiment, was het eveneens mogelijk om heterogene effecten te analyseren: welke personen maakten het meest gebruik van de voucher en in hoeverre resulteerde de voucher in betere uitkomsten voor de verschillende groepen personen?

Interventie: scholingsvoucher

Uitkomstmaten: besteding van de voucher

inkomen één jaar na ontvangst voucher

werkzaam één jaar na ontvangst voucher

training participatie één jaar na ontvangst voucher

Voorwaarden

De meest cruciale voorwaarde om een grootschalig gerandomiseerd experiment te laten slagen is dat de toewijzing van individuen of bedrijven aan de interventiegroep geheel willekeurig is. Daarnaast gelden nog een aantal voorwaarden die gelden voor alle andere interventiemethoden (waarbij interventie- en controlegroepen gedefinieerd worden):

- De interventie mag niet gerelateerd zijn aan de effectmaat. Als de verdeling in interventiegroepen en controlegroepen afhangt van de effectmaat, ontstaat er omgekeerde causaliteit.
- Selectieve respons (in het geval van vragenlijsten) moet uitgesloten kunnen worden.
- Spill-over effecten tussen de interventiegroep van de controlegroep mogen niet mogelijk zijn.
- Eventuele andere interventies moeten niet meer of minder betrekking hebben op de interventie- en controlegroep. Indien dit wel het geval is, is het onmogelijk om het verschil in de uitkomstvariabelen tussen de interventiegroep en de controlegroep toe te schrijven aan een bepaalde interventie.
- Men mag niet weten dat men onderdeel uitmaakt van een onderzoek naar de effectiviteit van de interventie. Het is daarom niet mogelijk om tevoren bekend te maken dat het experiment gaat plaatsvinden.

Mogelijkheden en beperkingen

De mogelijkheden bij een grootschalig gerandomiseerd experiment zijn talrijk. Als een koppeling aan de Enquête Beroepsbevolking of een ander representatief panel mogelijk is, is het mogelijk om een tal aan uitkomstvariabelen te analyseren. Ook is het mogelijk om de *dead-weight loss* te meten: in hoeverre had scholing ook zonder de stimulans

plaatsgevonden? Door het grootschalige karakter is het daarnaast mogelijk om heterogene effecten te schatten. Dit geeft inzicht in de mate waarin LLO-interventies een positief effect hebben voor specifieke (kwetsbare) groepen.

Toepasbaarheid van de evaluatiemethodiek

Deze methode is bij uitstek geschikt om overheidsinterventies te evalueren. Het kan hierbij gaan om het introduceren van scholingscheques gericht op alle werkenden in Nederland zoals het geval was in de studie van Schwerdt e.a. (2012). Hoewel in theorie de invoering van een individuele leerrekening ook via een grootschalig gerandomiseerd experiment geëvalueerd zou kunnen worden, kan het in de praktijk lastig worden om een dergelijke interventie van de grond te krijgen zonder ethische bezwaren.

Hoewel minder grootschalig, is het ook mogelijk om gerandomiseerde experimenten op sectoraal niveau op vergelijkbare wijze te evalueren.¹⁰ O&O fondsen kunnen willekeurig werknemers toewijzen die een scholingscheque krijgen en door middel van vragenlijsten onder alle werknemers inzichtelijk maken of er een hogere scholingsdeelname is onder de werknemers die een scholingscheque ontvangen hebben. Hierbij is men echter wel afhankelijk van de respons op de vragenlijsten.¹¹

Daarnaast lenen veldexperimenten zich ook uitstekend om onderzoek te doen naar verschillende communicatievormen omtrent LLO. Bijvoorbeeld bij het opstellen van digitale overzichten en portals kan worden geprobeerd om informatie steekproefsgevijs op verschillende manieren over te brengen in deze digitale omgevingen (zie ook o.a. alle initiatieven die worden ontwikkeld op het gebied van nudging) en te analyseren welke communicatievorm het meest effectief is.

Interessante publicaties

- Schwerdt, Messer, Woessmann & Wolter (2012). The impact of an adult education voucher program: Evidence from a randomized field experiment. *Journal of Public Economics*, Vol. 96, pp. 569-583
- Hidalgo, Oosterbeek & Webbing (2014). The impact of training vouchers on low-skilled workers, *Labour Economics*, Vol. 31, pp. 117-128

2.2 Trapsgewijs gerandomiseerd veldexperiment

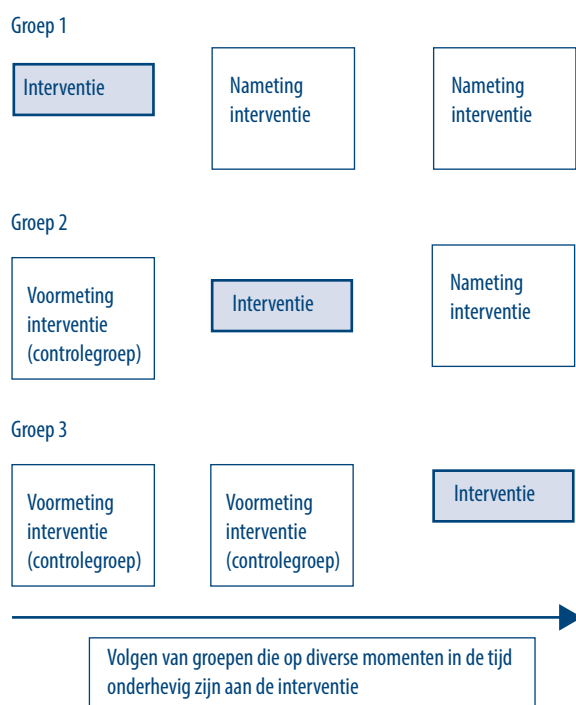
In de praktijk is een grootschalig gerandomiseerd experiment niet altijd mogelijk. Ethische kwesties spelen hierbij een belangrijke rol.¹² De onderverdeling in interventie- en controlegroepen zorgt er immers voor dat niet iedereen hetzelfde wordt behandeld. In de praktijk zijn er echter meer opties om toch op een gecontroleerde wijze, inclu-

¹⁰ Zie ook Hidalgo e.a. (2014).

¹¹ Hidalgo e.a. (2014) laten zien dat zelfs onder de werknemers die de vragenlijst hebben ingevuld, slechts 44% de vraag over hun inkomen heeft ingevuld. Selectieve non-respons kan leiden tot ruis in de effectmeting.

¹² In de gezondheidszorg waar gerandomiseerde experimenten de standaard zijn voor de introductie van nieuwe medicatie en behandelwijzen worden de ethische aspecten altijd getoetst door een ethische commissie die toestemming moet geven voor het mogen uitvoeren van een experiment.

sief randomisatieproces, een interventie te laten plaatsvinden. Het gaat dan om een trapsgewijze gerandomiseerde toewijzing van deelname op individueel-, afdelings-, of bedrijfsniveau. Het idee bij deze trapsgewijze gerandomiseerde toewijzing is dat iedereen uiteindelijk kan profiteren van de interventie, maar dat de interventie voor de controlegroep pas na de toetsing van het effect plaatsvindt. De groepsindeling wordt vervolgens gerandomiseerd, waarbij uitkomsten van mensen, afdelingen, of bedrijven die de interventie reeds hebben gehad vergeleken kunnen worden met die van groepen die de interventie nog moeten ondergaan. Een dergelijke vorm van randomisering is eenvoudiger te verantwoorden op ethische gronden en kan vaak ook vanuit praktische gronden eenvoudig worden verantwoord, omdat er sprake is van capaciteits- of financiële beperkingen. Bijvoorbeeld, als werknemers een gesprek willen hebben met een leerambassadeur, dan zal dit moeten worden ingepast in de agenda van de leerambassadeur. Niet iedereen kan immers op hetzelfde moment terecht. Een randomisering van de tijdslots kan het dan mogelijk maken om de interventiegroep (degenen die vroeg aan de beurt komen) van de controlegroep (degenen die laat aan de beurt komen) te onderscheiden.



Voorbeeldstudie

De Grip en Sauermann (2012) laten zien hoe waardevol een experiment kan zijn waarbij de tijdsallocatie van de toewijzing aan een interventie trapsgewijs is gerandomiseerd. In hun artikel identificeren zij de effecten van een werkgerelateerde training op de productiviteit van werknemers door gebruik te maken van een veldexperiment dat werknemers willekeurig toewijst over de tijd aan behandelings- en controlegroepen, gecombineerd met gegevens over de prestaties van de werknemers vóór en na de training. Uiteindelijk krijgt iedereen de werkgerelateerde training, maar het tijdstip verschilt dus steekproefsgewijs.

Interventie: trainingsdeelname

Uitkomstmaten: objectieve productiviteitsmaatstaven

Voorwaarden

Bij een trapsgewijs veldexperiment gelden de volgende voorwaarden:

- De randomisatie leidt tot een volledig willekeurige indeling in interventie- en controlegroepen.
- De groepen moeten over de tijd gevolgd kunnen worden.
- Omdat veldexperimenten in een minder gecontroleerde setting plaatsvinden dan laboratoriumexperimenten is er sprake van grotere standaarddeviaties, minder precisie en nauwkeurigheid. Veldexperimenten hebben daardoor een grotere steekproef nodig dan laboratoriumexperimenten. Dit geldt met name voor een trapsgewijs veldexperiment, omdat alle groepen op verschillende momenten in de tijd onderhevig zijn aan de interventie.
- Een laatste voorwaarde is dat de onderzoeker moet voorkomen dat het experimentele karakter van het onderzoek zelf een impact heeft op het gedrag van de verschillende groepen. Dit kan het beste worden vermeden door de volgorde van inloting als een praktisch noodzakelijk proces te presenteren. Zo maakten De Grip en Sauermann (2012) gebruik van de beperkte capaciteit van het opleidingscentrum om de trainingen te geven, waarbij de randomisatie van de loting als een eerlijk proces werd gezien.

Additioneel zijn er nog voorwaarden die voor alle evaluatiemethodes (die gebruik maken van interventie- en controlegroepen) gelden:

- De interventie mag niet gerelateerd zijn aan de effectmaat.
- Selectieve respons (in het geval van vragenlijsten) moet uitgesloten kunnen worden.
- Spill-over effecten tussen de interventiegroep van de controlegroep mogen niet mogelijk zijn.
- Eventuele andere interventies moeten niet meer of minder betrekking hebben op de interventie- en controlegroep.
- Men mag niet weten dat men onderdeel uitmaakt van een onderzoek naar de effectiviteit van de interventie.

Mogelijkheden en beperkingen

Een trapsgewijs veldexperiment heeft als groot voordeel dat onderzoekers en beleidsmakers meer controle hebben over de effectmeting en derhalve dus ook preciezer kunnen meten wat het effect is van LLO-interventies. Een ander groot voordeel is, zoals gezegd, dat lastige ethische kwesties die bij grootschalige gerandomiseerde veldexperimenten een rol kunnen spelen voor een groot deel kunnen worden voorkomen.

De grootste beperking van trapsgewijze veldexperimenten is dat de uitkomsten minder gemakkelijk repliceerbaar zijn. Veldexperimenten vereisen immers vaak speciale toegang tot de data. Ook zijn veldexperimenten vaak uitgevoerd in een specifieke context (bv. bedrijf of bedrijfssector), waardoor het onzeker is of ze ook opgaan in een andere context. Een tweede beperking is dat deze methode ook vereist dat er data verzameld worden voor meerdere meetmomenten. Een laatste beperking is dat de onderzoekers overeenstemming moeten bereiken met bedrijven of beleidsmakers om een dergelijke experimentele opzet te introduceren.

Toepasbaarheid van de evaluatiemethodiek

Deze techniek leent zich voor LLO-interventies op zowel individueel- als bedrijfsniveau. Op dit niveau kunnen onderzoekers direct onderhandelen met bedrijven of beleidsmakers om een gerandomiseerde trapsgewijze interventie te organiseren. Op sector- of regioniveau wordt dit een stuk lastiger, omdat er dan meerdere actoren betrokken moeten worden bij de goedkeuring en uitvoering van het experiment. Bovendien kan de betrokkenheid van meerdere actoren ook leiden tot minder controle op de juiste uitvoering van het experiment.

Een trapsgewijze allocatie kan onder andere worden ingezet bij de tijdsslots voor bezoek aan leerambassadeurs, het volgen van trainingen en cursussen, de verspreiding van informatie etc.

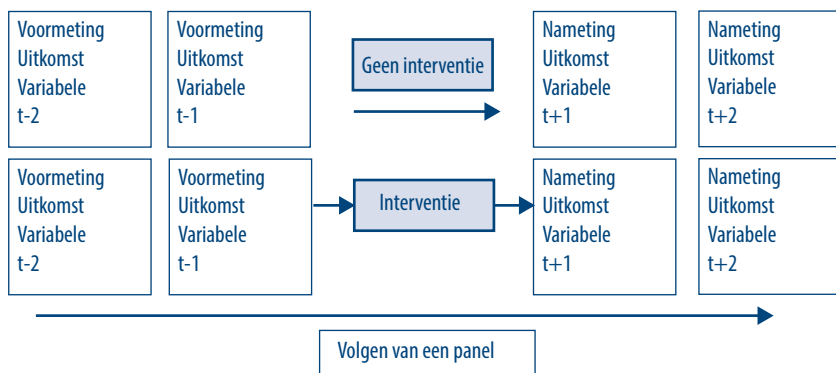
Interessante publicaties

- De Grip & Sauermann (2012). The effects of training on own and co-worker productivity: evidence from a field experiment. *The Economic Journal*, Vol. 122, pp. 376-399

2.3 Difference - in - Difference

Difference in Difference (DD) is een quasi experimentele techniek die probeert een experimenteel onderzoeksontwerp na te bootsen. Het meet het differentiële effect van een interventie op een 'interventiegroep' versus een 'controlegroep' in een natuurlijk experiment. Op basis van paneldata berekent de DD-methode het effect van een interventie op een uitkomst door te kijken of de uitkomstmaat voor de interventiegroep zich (gemiddeld) anders ontwikkelt dan voor de controlegroep. Deze methode vereist minimaal twee metingen voor de interventie en minimaal twee metingen na de interventie.

Een belangrijk kenmerk van deze methode is dat de toedeling van mensen aan de interventie of de controlegroep niet gerandomiseerd hoeft te zijn.



Voorbeeldstudie

Een interessant voorbeeld van een studie die een DD toepast is Görlitz (2010). In deze studie wordt het effect van een opleidingsvoucherprogramma op de investeringen van bedrijven in bijscholing in Duitsland geëvalueerd. Het voucherprogramma werd ingevoerd in de Duitse deelstaat Noord-Rijn-Westfalen en had als doel de trainingsdeelname van werknemers in kleine en middelgrote bedrijven te verhogen. Een voucher verlaagde de opleidingskosten met 50%. De DD-schatting die werd toegepast vergeleek de ontwikkeling van de gemiddelde trainingsparticipatie in bedrijven in de interventiegroep in Noord-Rijn-Westfalen met die van een controlegroep van bedrijven in andere deelstaten.

Interventie: scholingsvoucher

Uitkomstmaten: trainingsdeelname op bedrijfsniveau
trainingsintensiteit

Voorwaarden

Voor een effectmeting op basis van de DD-methode gelden de volgende voorwaarden:

- Bij de DD-aanpak moet er in ieder geval een inspectie van de trends voor de interventie van zowel de interventie- als controlegroepen plaatsvinden om te kijken of er voldaan wordt aan de 'common trend' assumptie. Dit betekent dat als er geen interventie zou plaatsvinden er geen verschillen tussen de interventiegroep en de controlegroep optreden. Als bijvoorbeeld, in de context van studie van Görlitz (2010), de trend in de trainingsdeelname in de Duitse deelstaten niet hetzelfde is, dan wordt het gemeten effect van de interventie immers vertekend. Er kan dan een interventie effect worden gemeten, terwijl het in feite gaat om een verschil in de ontwikkeling van de trainingsdeelname tussen de deelstaten dat er ook zonder de interventie zou zijn. Maar het kan ook dat afwijkende trends het effect van een interventie volledig verbloemen. Om te kijken of voldaan wordt aan de common-

trend assumptie moeten de uitkomstmaten op zoveel mogelijk momenten worden gemeten. Het minimumaantal voor- en nametingen is twee.

Daarnaast moet voldaan worden aan de voorwaarden die voor alle evaluatiemethoden (die gebruik maken van een interventie- en controlegroep) van toepassing zijn:

- De interventie mag niet gerelateerd zijn aan de effectmaat.
- Selectieve respons (in het geval van vragenlijsten) moet uitgesloten kunnen worden.
- Spill-over effecten tussen de interventiegroep van de controlegroep mogen niet mogelijk zijn. Bijvoorbeeld, de studie van Görlitz zou geen goede effectmeting opleveren als bedrijven die ook vestigingen hebben in andere deelstaten op grote schaal hun scholingsinvesteringen zouden verplaatsen naar Noord-Rijn-Westfalen om zo te profiteren van de lagere trainingskosten in deze deelstaat.
- Eventuele andere interventies moeten niet meer of minder betrekking hebben op de interventie- en controlegroep.
- Men mag niet weten dat men onderdeel uitmaakt van een onderzoek naar de effectiviteit van de interventie.

Mogelijkheden en beperkingen

De DD is een zeer sterk instrument als aan de bovenstaande voorwaarden wordt voldaan.

- Als aan al deze voorwaarden wordt voldaan kunnen de interventie effecten als causaal geïnterpreteerd worden.
- Een tweede voordeel is dat DD-analyses (net als een voor- en nameting) gemakkelijker op zowel individueel als organisatie-, sector, of regioniveau gebruikt kunnen worden dan een RD-design (zie sectie 2.4).
- Een additioneel voordeel ten opzichte van de RD-design is dat interventie- en controlegroep een ander startniveau van de uitkomstvariabele kunnen hebben. Een DD richt zich immers niet op de absolute niveaus, maar op een vergelijking in de veranderingen in de uitkomstvariabele voor en na de interventie.
- Een laatste voordeel is dat DD-analyses controleren voor veranderingen die door andere factoren dan de interventie worden veroorzaakt.

Naast het feit dat voldaan moet worden aan de genoemde voorwaarden, heeft de DD-methode als belangrijkste beperking dat er minimaal vier meetmomenten nodig zijn, terwijl de RD-methode maar twee meetmomenten vereist. Een DD kan daardoor uitsluitend met paneldata uitgevoerd worden.

Toepasbaarheid van de evaluatiemethodiek

De DD-methodiek leent zich voor effectmetingen van interventies op zowel individueel niveau- als op bedrijfs-, sector- of regioniveau. Bijvoorbeeld, door vergelijkingen te maken tussen de ontwikkeling in uitkomsten voor werknemers die in bedrijven werken waarin leerambassadeurs, individuele leerrekeningen, of andere interventies die tot

een leven lang ontwikkelen moeten aanzetten worden geïntroduceerd en de ontwikkeling van deze uitkosten voor werknemers die in bedrijven werken waarin dit niet het geval is. Hetzelfde kan gedaan worden op bedrijfsniveau of regioniveau. Als in bepaalde bedrijven, sectoren of regio's interventies plaatsvinden, dan kan de ontwikkeling van uitkomstmaatstaven op bedrijfs-, sector- of regioniveau worden vergeleken met een controlegroep van bedrijven, sectoren of regio's waar deze interventie niet plaatsvindt. Bijvoorbeeld, door het vergelijken van uittreding uit werkloosheidsuitkeringen tussen twee (of meer) regio's die op een ander moment een leerwerkloket opgezet hebben. Met behulp van een DD-analyse kan dan de effectiviteit worden onderzocht van de invoering van een leerwerkloket. In zoverre heeft de DD-analyse waarschijnlijk de grootste toepasbaarheid voor het meten van de effectiviteit van LLO-interventies. Belangrijk is echter wel dat voldaan wordt aan de genoemde basisvoorwaarden.

Interessante publicaties

- Görlitz, K. (2010). The effect of subsidizing continuous training investments—evidence from German establishment data. *Labour Economics*, 17(5), 789-798.

2.4 Regression discontinuity design

Een vierde manier om initiatieven rondom een 'Leven Lang Ontwikkelen' te evalueren is op basis van een regression discontinuity (RD) design. Een RD-design is een quasi experimenteel design dat effectief gebleken is in het causaal bepalen van de effectiviteit van interventies als uitsluitend een nameting voorhanden is.

Het meest eenvoudige RD-design lijkt sterk op een experimentele nameting van een interventie, waarbij een interventiegroep en een controlegroep die niet onderhevig is aan de interventie worden gevormd. Het verschil zit in de toekenning van de proefpersonen aan de twee condities. In het RD-design wordt er bijvoorbeeld op basis van een continue variabele een selectie gemaakt. Bijvoorbeeld als alleen mensen boven een bepaalde leeftijd in aanmerking komen voor de interventie. Ook kan de selectie gemaakt worden op basis van een testscore. Er wordt dan een grenswaarde vastgesteld die bepaalt of mensen tot de interventie- of de controlegroep behoren. Bijvoorbeeld als een leerrekening alleen gegeven wordt aan iedereen die ouder is dan 50, of aan iedereen die qua geletterdheid, of op een psychologische of vaardigheidstest onder een bepaalde grenswaarde scoort. Vervolgens wordt de impact van de interventie vergeleken door te focussen op degenen die net onder of net boven de grenswaarde zitten. Dit betekent dat een RD-design dus een lokaal interventie effect meet.



Voorbeeldstudie

In Leuven en Oosterbeek (2004) wordt een RD-design gebruikt om het effect te meten van een fiscale prikkel voor werkgevers om de trainingsparticipatie van hun medewerkers te vergroten. In 1998 heeft de Nederlandse overheid een nieuwe belastingwet ingevoerd die het mogelijk maakt om bedrijfsuitgaven voor werkgerelateerde opleidingen af te trekken van de belastingopgave. Bedrijven mochten bij hun belastingaangifte 40% van de opleidingskosten die worden gemaakt voor werknemers van 40 jaar of ouder aftrekken. Deze harde discontinuïteit in de opleidingskosten op basis van de leeftijd van de medewerkers maakte het mogelijk om de effecten van de belastingaftrek op de trainingsdeelname te identificeren. De resultaten van deze studie toonden aan dat het trainingsdeelname van werknemers net boven de leeftijd van 40 jaar significant hoger is dan de deelname van werknemers net onder de leeftijd van 40 jaar. Een snelle vergelijking zou dus tot de conclusie kunnen leiden dat het verlagen van opleidingskosten een positief effect heeft op de trainingsdeelname van werknemers. De studie liet echter ook zien dat dit verschil vooral het gevolg is van het feit dat werkgevers trainingsinvesteringen uitstellen tot een werknemer 40 jaar is geworden waardoor er recht is op belastingaftrek. Naast de nameting hadden de auteurs ook toegang tot een voormeting waaruit bleek dat de trainingsdeelname onder mensen net onder de leeftijd van 40 jaar was gedaald. Dit betekent dat de hogere trainingsdeelname van werkgevers net boven de 40 niet geïnterpreteerd mag worden als een stimulerend effect van de nieuwe belastingwet. Dit verklaart ook waarom Leuven en Oosterbeek (2004) geen verdere effecten vonden van de verandering van de belastingaftrek op de loonuitkomsten van 40-plussers.

Interventie: Belastingwet 1998
Uitkomstmaten: trainingsdeelname
loonuitkomsten

Voorwaarden

Een RD-design kan onder bepaalde voorwaarden voldoende zijn om te kunnen spreken over een effectmeting. Daarbij zijn de volgende zaken van belang:

- Een RD-design vereist dat de toedeling van mensen aan de interventie- en de controlegroep zo goed als steekproefsgewijs verloopt. Dit betekent dat goed moet worden gecheckt of de interventie- en de controlegroep op observeerbare kenmerken identiek zijn.
- Niemand mag in staat zijn om de variabele op basis waarvan de interventiegroep van de controlegroep wordt onderscheiden te manipuleren, zodat zij zichzelf kunnen selecteren in de interventie- of controlegroep.
- De variabele op basis waarvan het onderscheid wordt gemaakt tussen de interventie- en de controlegroep moet een continue variabele zijn, zodat de grenswaarde zo precies mogelijk kan worden bepaald. Discontinuïteit is vereist voor het onderscheiden van de interventie- en de controlegroep.
- Voor een RD-design is het van belang om een optimale bandbreedte rondom de grenswaarde op basis waarvan de interventiegroep van de controlegroep wordt

onderscheiden te bepalen. Het RD-design berekent immers een lokaal interventie-effect precies rondom de grenswaarde. Een te brede bandbreedte kan ertoe leiden dat er te grote verschillen ontstaan tussen de interventie- en de controlegroep, of andere interventies de meting van de effectiviteit van interventie waarin men geïnteresseerd is verstoren. Hiertoe is een relatief grote N vereist.

- Bij een RD-design moet altijd gecontroleerd worden voor mogelijke trends in de variabele op basis waarvan de interventiegroep van de controlegroep wordt onderscheiden. Een effectmeting kan anders namelijk het gevolg zijn van deze trends in plaats van de trendbreuk rondom de grenswaarde.

Daarnaast spelen de voorwaarden die voor alle interventiemethodes (die gebruik maken van een interventie- en controlegroep) een rol:

- De interventie mag niet gerelateerd zijn aan de effectmaat.
- Selectieve respons (in het geval van vragenlijsten) moet uitgesloten kunnen worden.
- Spill-over effecten tussen de interventiegroep van de controlegroep mogen niet mogelijk zijn.
- Eventuele andere interventies moeten niet meer of minder betrekking hebben op de interventie- en controlegroep.
- Men mag niet weten dat men onderdeel uitmaakt van een onderzoek naar de effectiviteit van de interventie.

Mogelijkheden en beperkingen

Het grootste voordeel van een RD-design is dat er in principe alleen een nameting vereist is na een interventie die een sterke discontinuïteit veroorzaakt. Het is dan mogelijk om op een betrouwbare manier een causaal interventie effect te meten, zelfs als er niet voldoende tijd is om een voormeting te introduceren. Bij een goede uitvoering van het RD-design waarin aan alle bovenstaande voorwaarden is voldaan, zullen effectmetingen in een nameting niet significant verschillen van die in een situatie waarin zowel een voor- als nameting beschikbaar is.

De methode moet echter aan een aantal duidelijke eisen voldoen om het effect van een interventie goed te kunnen meten. Ten eerste vereist het kunnen meten van een interventie-effect dat er voldoende observaties zijn rondom de grenswaarde. In het geval van de evaluatiestudie van Leuven en Oosterbeek (2004) betekent dit dat er voldoende waarnemingen moeten zijn onder werknemers die nog net geen 40 jaar zijn en die net 40 jaar geworden zijn. Dit betekent doorgaans dat RD-designs alleen mogelijk zijn als er een grote datasets is, of als datasets specifiek ontworpen worden om voldoende mensen te analyseren die net boven of onder de grenswaarde zitten.¹³ Ten tweede mogen er geen andere interventies of ontwikkelingen tegelijkertijd plaatsvinden. Ten

¹³ Een goed voorbeeld daarvan is het onderzoek van De Grip e.a. (2012) onder mannen die in 1949 of 1950 waren geboren naar het effect van het versoberen van de pensioenrechten voor iedereen die na 1949 is geboren.

derde mag de toewijzing van mensen aan de interventie- of de controlegroep niet door zelfselectie of andere factoren beïnvloed worden. Dit betekent dat onderzoekers uitgebreid moeten testen op observeerbare variabelen om aan te tonen dat deze groepen aan elkaar gelijk zijn. Datasets moeten bij voorkeur dus veel data bevatten over diverse onderwerpen. Echter, zelfs dan kan nog niet geheel uitgesloten worden dat deze groepen op niet-observeerbare categorieën niet van elkaar verschillen. Ook als aan al deze eisen is voldaan, heeft een RD-design nog nadelen. Zoals de resultaten van Leuven en Oosterbeek (2004) laten zien, zijn *crowding out* processen met gebruik van één nameting moeilijk te achterhalen. Crowding out treedt op als een positief resultaat van een interventie ten koste is gegaan van iemand anders. Een voormeting kan daarbij helpen om crowding out processen te identificeren. Ook hebben RD-designs het nadeel dat ze slechts een lokaal interventie-effect schatten. Hoewel de resultaten intern valide kunnen zijn, is de vraag in welke mate deze ook extern valide zijn. In het geval van de studie van Leuven en Oosterbeek (2004) is het bijvoorbeeld de vraag in hoeverre de resultaten hetzelfde geweest zouden zijn als de leeftijdsgrens voor belastingaftrek bij 30 of bij 50 jaar was gelegd. Voorzichtigheid is derhalve geboden bij de interpretatie van de resultaten van RD-designs en het is belangrijk om oog te blijven hebben voor de specifieke context waarin het onderzoek heeft plaatsgevonden.

Toepasbaarheid van de evaluatiemethodiek

Aangezien een RD-design een voldoende kritische massa vereist rondom de grenswaarde om het lokale interventie-effect te kunnen meten, is deze techniek vooral toepasbaar in grote steekproeven of administratieve datasets, die doorgaans maar een beperkte set van uitkomstvariabelen hebben. RD-designs lenen zich daarom vooral voor het meten van uitkomsten op individueel niveau. Daarentegen zijn RD-designs minder geschikt voor het meten van de effecten van een interventie op sectorniveau, regioniveau, of op bedrijfsniveau. Derhalve is het aantal typen interventies dat door een nameting geëvalueerd kan worden enigszins beperkt. Een goed voorbeeld van een LLO-interventie die door middel van een RD-design kan worden geëvalueerd is het aanbod van scholingscheques in grote bedrijven. Het kan daarbij zowel gaan om functiegerelateerde als employability-gerelateerde vouchers (zoals de E-miles van Philips). Een belangrijke voorwaarde blijft echter dat er een sterke discontinuïteit moet zijn in wie wel of geen cheque krijgt. Bijvoorbeeld, het aanbod van scholingscheques zou kunnen afhangen van een specifieke leeftijd, individuele productiviteitsmetingen of een grensscore van een eerder gemaakte motivatie-, psychologische-, cognitieve- of vaardigheidstest, waarvan werknemers niet wisten dat deze later gebruikt zouden worden voor de toekenning van de cheques.

Bij het meten van de effecten van de invoering van een individuele leerrekening zal het lastig zijn om een discontinuïteit in te bouwen. Wel kan gedacht worden aan RD-designs met betrekking tot de informatievoorziening over deze leerrekeningen. Bijvoorbeeld als werknemers die ouder zijn dan 50 jaar extra informatie krijgen over een goede manier om de individuele leerrekening te gebruiken of over het belang van het leven lang ontwikkelen voor hun duurzame inzetbaarheid. Ook is het mogelijk om de

hoogte van de leerrekening afhankelijk te maken van een bepaalde testscore of leeftijd. Bijvoorbeeld, als het budget van de leerrekening hoger wordt als werknemers 45 jaar oud worden, wat een cruciale leeftijd is waarop de employability substantieel begint af te nemen (Montizaan & Fouarge, 2016).

Interessante publicaties

- Leuven, E., & Oosterbeek, H. (2004). Evaluating the effect of tax deductions on training. *Journal of Labor Economics*, Vol. 22(2), pp. 461-488.
- CPB (2016). Evaluatie aftrekpost scholingsuitgaven. CPB Notitie, 1 juli 2016 <https://www.cpb.nl/sites/default/files/omnidownload/CPB-Notitie-1juli2016-Evaluatie-aftrekpost-scholingsuitgaven.pdf>

2.5 Voor- en nameting

Een laatste manier om LLO-initiatieven te evalueren is op basis van een voor- en nameting. Hierbij moet echter opgemerkt worden dat er met deze methode geen causale effecten kunnen worden aangetoond, omdat er geen controlegroep aanwezig is. Bij een voor- en nameting worden de uitkomstmaten simpelweg voor en na de invoering van een interventie gemeten. Deze methode zou daarom alleen in uitzonderlijke gevallen gehanteerd moeten worden als de te onderzoeken interventie dermate belangrijk is dat een *verkennend* onderzoek naar de werking ervan gerechtvaardigd is en het niet mogelijk is om een van de andere methoden toe te passen.



Voorbeeldstudie

In Gerards e.a. (2014) wordt een voor- en nameting gebruikt om de effecten van de invoering van een scholingsvoucher te analyseren. Het betreft een scholingsvoucher binnen Philips gericht op het vergroten van de (interne en externe) inzetbaarheid van de werknemers. Op 1 januari 2009 kregen alle werknemers van Philips 1000 zogenaamde E-miles (Employability-miles). Deze E-miles konden gebruikt worden voor deelname in een beperkt aantal cursussen gericht op het vergroten van de (bewustwording van de) inzetbaarheid. De E-miles waren geldig tot en met 31 december 2009. Door het afnemen van een tweetal vragenlijsten onder een gerandomiseerde groep werknemers werd inzicht gegeven in de mate waarin een specifieke scholingsvoucher gerelateerd is aan op employability gerelateerde cursusdeelname. De eerste vragenlijst werd in december 2008, vlak voor de invoering van de E-miles, afgenomen en de tweede vragenlijst in februari 2010 (net na het verlopen van de E-miles). Beide vragenlijsten bevatten vragen over leer gerelateerde variabelen, persoonlijkheid en de werkplek (o.a. gebruik van HR-instrumenten). In de tweede vragenlijst werd additioneel gevraagd naar

het vouchergebruik. Ongeveer een kwart van de aangeschreven mensen hebben beide vragenlijsten ingevuld.

Interventie: scholingsvoucher
Uitkomstmaten: besteding van de voucher
totale cursusdeelname
gepercipieerde inzetbaarheid
gepercipieerde kans op baanverlies

Voorwaarden

Een voor- en nameting kan slechts onder een scala aan zeer strikte voorwaarden voldoende zijn om te kunnen spreken over een onderzoek naar de werking van LLO-interventies. Hierbij zijn de volgende zaken van belang:

- De voormeting moet plaatsvinden voordat de betrokkenen op de hoogte zijn van de geplande interventie. Dit is met name van belang als de uitkomstvariabelen subjectief van aard zijn zoals bijvoorbeeld gepercipieerde inzetbaarheid.
- Er mag slechts één interventie tegelijkertijd plaatsvinden. Indien dit niet het geval is, is het onmogelijk om het verschil in de uitkomstvariabelen tussen de voor- en nameting toe te schrijven aan een bepaalde interventie.
- Er mogen tussen de voor- en nameting geen andere voor de uitkomstvariabelen relevante ontwikkelingen plaatsvinden. Als er tegelijkertijd ook sprake is van bijvoorbeeld de introductie van een persoonlijk ontwikkelingsgesprek, dan kan het effect van de interventie niet geïdentificeerd worden.
- Er mag in de uitkomstvariabelen geen ontwikkeling in de tijd zijn. Immers, als er bijvoorbeeld al enige tijd sprake is van een toenemende deelname aan cursussen of een stijging in de bewustwording van het belang van een goede inzetbaarheid, dan kan onterecht een effect toegeschreven worden aan de interventie.

Daarnaast spelen de volgende (voor elke besproken evaluatiemethode geldende) voorwaarden:

- De interventie mag niet gerelateerd zijn aan de effectmaat.
- Selectieve respons (in het geval van vragenlijsten) moet uitgesloten kunnen worden. Vooral het risico op uitdiensttreding als gevolg van de interventie (zoals het geval bij de hierboven beschreven E-miles) vormt een gevaar voor de identificatie van effecten.¹⁴
- Men mag niet weten dat men onderdeel uitmaakt van een onderzoek naar de effectiviteit van de interventie.

¹⁴ Op basis van het aantal *rebounces* op de uitnodigingsmail van de tweede vragenlijst concluderen Gerards e.a. (2014) dat werknemers die vertrokken zijn bij Philips geen gevaar vormen voor de effectmeting.

Mogelijkheden en beperkingen

Het grootste voordeel van een effectmeting op basis van een voor- en nameting is dat een interventie tegelijkertijd plaats kan vinden voor een groot aantal mensen of bedrijven.¹⁵ Er zijn dus geen gerandomiseerde treatment- en controlegroepen nodig. Het gebruik van vragenlijsten maakt het bovendien mogelijk om meer informatie boven tafel te krijgen dan alleen de grootte van het effect. Daarnaast kan:

1. inzicht verkregen worden in de mate waarin verschillende groepen werkenden (o.b.v. persoonskenmerken, persoonlijkheid, baankenmerken) gebruik maken van de voucher (*heterogene effecten*);
2. inzicht verkregen worden in de factoren die positief of negatief gerelateerd zijn aan het gebruik maken van de voucher (*mechanismen*);
3. geanalyseerd worden of de voucher tot een toename van het aantal gevolgde cursussen, of dat er alleen sprake is geweest van een substitutie van cursussen¹⁶ (*crowding-out effect*).

De voorwaarden dat er geen andere interventies of ontwikkelingen parallel mogen lopen en dat er geen trend mag zijn in de uitkomstvariabelen vormen een groot nadeel van deze methode. Het is immers niet altijd vooraf bekend welke ontwikkelingen zich zullen voordoen en ook bestaande trends zijn niet altijd bekend. Alleen al het besluit om een scholingsimpuls te geven, reflecteert vaak dat het management van het bedrijf gericht is op het vergroten van het leergedrag binnen het bedrijf. In dat geval is het niet mogelijk om het effect van de bewuste interventie te isoleren van de algehele leer-cultuur en managementfocus op de cursusdeelname. Hierdoor kunnen de gevonden 'effecten' ofwel een te positief ofwel een te negatief beeld van de feitelijke 'causale' effecten van de interventie.

Toepasbaarheid van de evaluatiemethodiek

In de context van LLO is een voor- en nameting een relatief eenvoudige manier op interventies op bedrijfsniveau te evalueren. Er is op bedrijfsniveau immers een grotere kans dat aan de bovengenoemde voorwaarden voldaan wordt dan op sector- en regioniveau. Op sector- en regioniveau spelen immers meer trends en ontwikkelingen die buiten de regie van de initiatiefnemer van de interventie liggen. Bij een interventie op bedrijfsniveau kan gekeken worden naar effecten op zowel individueel als bedrijfsniveau.

Er zijn verschillende typen interventies die door een voor- en nameting geëvalueerd kunnen worden. Allereerst kan gedacht worden aan *scholingsvouchers*. Het kan daarbij zowel gaan om functiegerelateerde als employability gerelateerde vouchers (zoals de E-miles van Philips).

Het invoeren van *leerambassadeurs* binnen een bedrijf is ook een interventie waarbij, zolang aan de bovengenoemde voorwaarden voldaan wordt, een voor- en nameting gebruikt zou kunnen voor een verkennend onderzoek naar de werking ervan. Ook hier

¹⁵ Dit is geen vereiste. Een voor- en nameting kan ook gebruikt worden bij een kleinere onderzoekspopulatie.

¹⁶ Zie hiervoor ook o.a. Gerards e.a. (2014)

is het bijvoorbeeld mogelijk om te analyseren of er voor en na de invoering van de leerambassadeur(s) meer scholing gevolgd wordt dan daarvoor. Daarnaast is het interessant om te kijken of de bereidheid van werknemers om aan scholing deel te nemen en de bewustwording van de noodzaak tot verdere ontplooiing door het invoeren van leerambassadeurs is toegenomen. In Hoofdstuk 3 gaan we verder in op de mogelijkheden die er zijn om daadwerkelijk te kunnen spreken van effectmetingen rondom de invoering van leerambassadeurs.

Ook zijn er mogelijkheden om de effecten van ondersteuning voor de regionale infrastructuur voor LLO door middel van een voor- en nameting te meten. Zo kan er gekeken worden naar het succesvol plaatsen van werkzoekenden voor én na het ontstaan van leerwerkloketten.¹⁷

Interessante publicaties

- Gerards, De Grip & Witlox (2014). "Employability-miles" and worker employability awareness. *Applied Economics*, Vol. 46, pp. 952-965

2.6 Rangschikking interventiemethoden

Het gebruik van interventiemethoden zal doorgaans afhangen van hoe de interventie is vormgegeven en welke databronnen beschikbaar zijn. Maar als onderzoekers meerdere opties tot hun beschikking hebben, is het belangrijk om een goed beeld te hebben van welke methode de voorkeur verdient. Tabel 2.1 rangschikt de methoden op basis van de mate van interne en externe validiteit op basis van de voorwaarden en beperkingen die in het voorafgaande zijn besproken. Daarnaast wordt ook de Maryland schaal meegenomen in de ranking. De zwakste interventiemethode is overduidelijk de voor-en nameting. Doordat er geen controlegroep is, is het onmogelijk om het causale effect van een interventie te meten. Voor beleidsmakers zijn de onderzoeksresultaten van deze methode derhalve maar beperkt bruikbaar. Een RD zonder voormeting verdient derhalve de voorkeur boven een voor- en nameting, omdat bij de RD-methode wel interventie- en controlegroepen worden gecreëerd. Echter doordat deze evaluatiemethode alleen lokale interventie effecten meet, hebben de resultaten alleen betrekking op de mensen of bedrijven direct rondom de interventiegrens.

Een DD verdient daarom de voorkeur boven een RD zonder voormeting. De reden hiervoor is dat een DD niet slechts lokale interventie-effecten meet, waardoor de externe validiteit van de effectmeting veel groter is. Nadeel is wel dat bij de DD-methode het onderscheid tussen interventiegroepen en controlegroepen niet gerandomiseerd is. Een RD met voormeting kan in theorie daarom nog net iets sterker zijn. Immers wordt gebruik gemaakt van zowel interactie- en controlegroepen die zo goed als gerandomiseerd zijn als van een voor- en nameting. Een RD met voormeting combineert daarmee

¹⁷ Een difference in difference onderzoek is echter een betere evaluatiemethode. Deze methode vereist echter meer meetmomenten en zowel een treatment als een controlegroep. Zie sectie 2.3.

de voordelen van beide methoden. Een trapsgewijze gerandomiseerde toewijzing van deelname op individueel- of afdelingsniveau of een grootschalig veldexperiment staan aan de top van de ranking. De reden hiervoor is dat onderzoekers bij deze methoden in staat zijn om de hele interventie te controleren, waardoor de meeste problemen die zich voordoen bij de andere methoden zich niet zullen voordoen. De interne validiteit van deze methoden is derhalve zeer groot. De externe validiteit kan echter iets minder groot zijn dan bij een RD met voormeting of een DD door de specifieke setting van het experiment.

TABEL 2.1 Ranking van evaluatiemethoden op basis van de Maryland schaal en de mate van interne en externe validiteit

Methode	Maryland schaal	Mate van interne validiteit	Mate van externe validiteit
Grootschalig gerandomiseerd experiment	5	Zeer grote interne validiteit door randomisatie	Redelijke tot grote externe validiteit afhankelijk van de specifieke setting waarin experiment plaatsvindt (meestal binnen bedrijven / organisaties)
Trapsgewijs gerandomiseerd veldexperiment	5	Zeer grote interne validiteit door randomisatie	Redelijke tot grote externe validiteit afhankelijk van de specifieke setting waarin experiment plaatsvindt (meestal binnen bedrijven / organisaties)
Regression discontinuity design met voormeting	4	Zeer grote interne validiteit door quasi-randomisatie. De voormeting zorgt voor de identificatie van anticipatie-effecten	Redelijke tot grote externe validiteit door het meten van een lokaal interventie effect
Difference - in - difference methode	3	Grote interne validiteit. Hangt vooral cruciaal af van de common trend assumptie	Grote externe validiteit
Regression discontinuity design zonder voormeting	4	Matige interne validiteit. Deze hangt cruciaal af van het gebrek aan anticipatie effecten (die door een gebrek aan een voormeting niet kunnen worden gemeten)	Redelijke externe validiteit door het meten van een lokaal interventie effect
Voor- en nameting	1/2	Geen vergelijking tussen interventie- en controlegroepen. Interne validiteit is zwak	Beperkt, vanwege selectiviteit en de impact van andere interventies en trends waarmee deze techniek geen rekening mee houdt

3 FOCUS OP INDIVIDUELE LEERREKENINGEN EN LEERAMBASSADEURS

In dit hoofdstuk laten we zien hoe vanuit het in hoofdstuk 2 in kaart gebrachte evaluatieraamwerk bekeken kan worden op welke manieren LLO-interventies geëvalueerd kunnen worden. Daarbij richten we ons op twee specifieke interventies.

3.1 Individuele leerrekeningen

Het Nederlandse kabinet zet zich in voor een doorbraak op het gebied van LLO en een sterke en positieve leercultuur. In het regeerakkoord is afgesproken dat het kabinet in dit kader wil bevorderen dat iedereen een individueel budget voor algemene scholing en ontwikkeling (leerrekening) tot zijn of haar beschikking krijgt. Het idee achter individuele leerrekeningen is dat iedereen zo meer regie krijgt over zijn of haar loopbaan en zich beter kan aanpassen aan veranderingen op de arbeidsmarkt. Werknemers krijgen daarmee meer eigen verantwoordelijkheid voor de ontwikkeling van hun competentieprofiel en duurzame inzetbaarheid op de arbeidsmarkt.

Er zijn twee manieren waarop het kabinet dit wil gaan stimuleren:

1. via een fiscale regeling (verbeterde fiscale facilitering?) voor private individuele leerrekeningen
2. via een uitgavenregeling voor publieke individuele leer- en ontwikkelbudgetten, die naar verwachting in 2020 kan worden ingevoerd.

Private leerrekeningen

In sommige bedrijven en sectoren zijn er al individuele leer- en ontwikkelrekeningen, in de vorm van leerrekeningen, vouchers of trekkingsrechten. Deze zijn allemaal bedoeld om de werknemer in het bedrijf of de sector meer regie te geven op het terrein van hun ontwikkeling. Een goed voorbeeld hiervan is de James Leerrekening. Dit is een persoonlijk ontwikkelbudget dat men kan besteden aan een opleiding, training en loopbaanbegeleiding. Over dit budget hoeft men geen loon- of inkomstenbelasting te betalen en werknemers kunnen het meenemen als ze een nieuwe baan krijgen. Werkgevers kunnen het budget van hun medewerkers vullen, maar werknemers kunnen zelf ook geld inleggen op hun leerrekening, of een storting op hun rekening uitruilen tegen andere arbeidsvoorwaarden.

Hoewel de leerrekening al in een aantal gevallen wordt toegepast, wordt de groot-scheepse introductie hiervan gehinderd door onduidelijkheden over de fiscale behandeling van private leerrekeningen. Het betreft dan specifiek de gerichte vrijstelling in de loonheffingen en de meeneembaarheid ofwel portabiliteit van individuele leer- en ontwikkelbudgetten als iemand ergens anders gaat werken.

Een ander belangrijk issue is dat een succesvolle introductie van private leerrekeningen waarschijnlijk vereist dat er andere investeringen in het LLO plaatsvinden. Een belangrijke voorwaarde voor het succes ervan ligt immers in de mate waarin iemand die zo'n leerrekening krijgt gemotiveerd is om zich verder te ontwikkelen en ook een goed beeld heeft van de voordelen van het volgen van een bepaalde opleiding. Zo zou de introductie van de James leerrekeningen gecombineerd kunnen worden met vormen van loopbaancoaching en het aanbieden van een digitale omgeving waarin iemand kan zoeken naar informatie over cursussen en trainingen en baankansen, job crafting en coaching. Het spreekt voor zich dat als het succes van een interventie mede bepaald wordt door complementaire interventies, het moeilijker wordt om de effecten van de leerontwikkeling te isoleren.

Publiek individueel leer- en ontwikkelbudget

Het kabinet probeert daarnaast een uitgavenregeling te introduceren om publieke middelen te storten op een individueel leer- of ontwikkelbudget. De middelen voor deze uitgavenregeling komen uit de voorgenomen afschaf van de fiscale aftrek voor scholingsuitgaven (€ 218 miljoen). Het kabinet wil deze middelen gericht inzetten op mensen die zich moeten bij- of omscholen naar een beroep of sector waarin zij betere baankansen hebben.

Effectmetingen van leerrekeningen

De effectiviteit van een leerrekening kan het beste gemeten worden met een DD-design. Een voor- en nameting, RD-design en een (trapsgewijs) experiment zijn alternatieven.

Een RD-design leent zich in beperkte mate voor het evalueren van private en publieke leerrekeningen. In beide gevallen zullen deze leerrekeningen immers aan grote groepen werknemers beschikbaar worden gesteld (al dan niet binnen bedrijven, sectoren of heel Nederland) waardoor het vaak moeilijk zal zijn om een onderscheid te maken tussen interventie- en controlegroepen op basis van specifieke grenswaarden die een discontinuïteit creëren in het al dan niet toegang hebben tot een leerrekening. Toch is het niet helemaal onmogelijk om een RD-design te creëren, aangezien het kabinet zelf heeft aangegeven dat het beperkte budget ertoe zal leiden dat zij de middelen specifiek zullen moeten gaan inzetten. Om *dead-weight loss* te voorkomen, zou het goed zijn om de leerrekening te introduceren voor groepen werkenden die relatief weinig scholing volgen, zoals lager opgeleiden, flexwerkers en oudere medewerkers.

Bij het afbakenen van deze groepen zou kunnen worden gewerkt met een grenswaarde die bepaalt of men al dan niet tot een risicogroep voor een te lage scholingsdeelname behoort. Mogelijke voorbeelden van een dergelijke grenswaarde zouden kunnen zijn:

1. Een specifieke leeftijdsgrens, aangezien oudere werknemers die hun werk kwijtraken slechtere perspectieven hebben op een nieuwe baan.
2. De door het ROA ontwikkelde Indicator Toekomstig Arbeidsmarktperspectief (ITA) die aangeeft welke vraag-aanbod-discrepancie er voor iemand met een bepaalde opleidingsachtergrond de komende zes jaar te verwachten is. Als het arbeidsaanbod kleiner is dan de vraag, en de ITA dus kleiner dan of gelijk is aan 1,00, dan wordt het arbeidsmarktperspectief als goed getypeerd. Als de waarde van de ITA zelfs kleiner dan of gelijk is aan 0,85, dan wordt gesproken van een zeer goed arbeidsmarktperspectief. Daarentegen, als de ITA een waarde heeft tussen de 1,05 en 1,15 en er dus een aanbodoverschot wordt verwacht, dan wordt het arbeidsmarktperspectief voor het desbetreffende opleidingstype als matig aangeduid en bij een ITA groter dan 1,15, als slecht. Deze grenzen bieden goede mogelijkheden om risicogroepen te definiëren en daar de leerrekening voor beschikbaar te stellen. Om de effecten van deze interventie goed te kunnen meten, moeten groepen werknemers worden gemonitord die op basis van deze ITA-score net niet of juist net wel in aanmerking komen voor een leerrekening.

Ook bij private leerrekeningen zijn er mogelijkheden om een RD-design te implementeren. Bijvoorbeeld bij het aanbod van scholingscheques in grote bedrijven of extra investeringen in de leerrekening door de werkgever. Individuele werkgevers zouden discontinuïteiten kunnen inbouwen bij de toekenning van de scholingscheque of de hoogte van de cheques op basis van:

1. Een specifieke leeftijd.
2. Individuele productiviteit of hoogte salaris c.q. salarisschaal.
3. Een grensscore van een eerder gemaakte motivatie-, psychologische-, cognitie- of vaardigheidstest.

Een andere mogelijkheid die overwogen kan worden is dat een RD-design wordt geïntroduceerd in de communicatie rondom de leerrekening. Hoewel iedereen in principe recht heeft op een leerrekening (publiek of privaat), kan bijvoorbeeld bij een specifieke groep aandacht worden besteed aan de communicatie richting specifieke risicogroepen om ervoor te zorgen dat deze meer budget gaan storten in hun rekening.

RD-designs hebben echter ook nadelen (zie hoofdstuk 2). Zo kunnen leeftijdsgrenzen leiden tot crowding out processen in de scholingsdeelname of andere investeringen in iemands menselijk kapitaal. Zo kunnen mensen hun trainingsdeelname uitstellen als het experiment het aantrekkelijk maakt om boven een bepaalde leeftijd een training te gaan volgen. Daarom is het aan te bevelen om in ieder geval ook één voormeting te hebben zodat deze processen geïdentificeerd kunnen worden. Daarnaast zijn er grote steekproeven nodig om een lokaal interventie-effect te meten. Dit laatste kan vooral

een groot probleem opleveren bij het meten van de effecten van leerrekeningen die maar bij de interventiegroep in één bedrijf zijn uitgezet.

Ten opzichte van een RD-design lijkt een voor- en een nameting een eenvoudigere manier om het effect van de introductie van leerrekeningen te meten. De scholingsdeelname voor en na de introductie van leerrekeningen kan immers vrij eenvoudig vergeleken worden. Het simpelweg vergelijken hiervan kan echter een te positief beeld opleveren. Bijvoorbeeld, als de economische groei ertoe heeft geleid dat er los van de introductie van de leerrekening meer opleidingsbudget beschikbaar komt.

Een DD-methode verdient echter de voorkeur ten opzichte van een voor- en nameting. DD-analyses controleren immers voor veranderingen die door andere factoren dan de leerrekening worden veroorzaakt. Naast het evalueren van de introductie van leerrekeningen kunnen ook specifieke aspecten van een leerrekening geëvalueerd worden met een DD-methode. Als iedereen (binnen een bedrijf of sector) een leerrekening tot zijn beschikking krijgt, dan kan er bijvoorbeeld geëxperimenteerd worden door voor één groep een houdbaarheidsdatum aan de leerrekening gekoppeld kunnen worden.

Een trapsgewijs gerandomiseerd veldexperiment en een puur experiment zijn minder geschikt voor het meten van de effecten van de introductie van een leerrekening. De invoering van een private leerrekening hangt vooralsnog sterk af van de CAO-onderhandelingen en de bereidheid van individuele werkgevers om mee te werken. Er zullen dus meerdere actoren betrokken moeten worden bij de uitvoering van een experiment. Daarnaast zijn de leerrekeningen doorgaans beschikbaar voor iedereen in een specifieke sector nadat hierover in de CAO overeenstemming is bereikt.

Binnen bedrijven zou het echter wellicht nog mogelijk zijn dat werkgevers budgetten steekproefsgewijs over verschillende perioden alloceren aan hun werknemers. Ook de overheid zou er bij de invoering van de publieke leerrekening voor kunnen kiezen om steekproefsgewijs interventie- en controlegroepen te definiëren. De overheid zou daarnaast, vanwege beperkte middelen, kunnen besluiten om het individueel budget op de publieke leerrekening te variëren op basis van de regionale arbeidsmarktkansen binnen Nederland, of om de publieke leerrekening in eerste instantie uitsluitend te introduceren in regio's met de slechtste arbeidskansen.

Tabel 3.1 vat de mogelijkheden voor het evalueren van leerrekeningen samen.

TABEL 3.1 Overzicht gebruik evaluatiekader voor private en publieke leerrekeningen

	Toepasbaarheid	Niveaus	Mogelijke obstakels
Grootschalig gerandomiseerd experiment	Beperkt	<ul style="list-style-type: none"> • Individueel niveau • Bedrijfsniveau 	<ul style="list-style-type: none"> • Mogelijke ethische kwesties
Trapsgewijs gerandomiseerd veldexperiment	Beperkt	<ul style="list-style-type: none"> • Bedrijfsniveau 	<ul style="list-style-type: none"> • Alleen mogelijk met gebruik van paneldata
DD-methode	Zeer goed	<ul style="list-style-type: none"> • Bedrijfsniveau • Sectorniveau • Regionaal niveau 	<ul style="list-style-type: none"> • Alleen mogelijk met gebruik van paneldata • Common trend assumptie nodig • Minimaal twee voormetingen nodig
RD-design met voormeting	Beperkt	<ul style="list-style-type: none"> • Individueel niveau 	<ul style="list-style-type: none"> • Crowding out effecten
RD-design zonder voormeting	Beperkt	<ul style="list-style-type: none"> • Individueel niveau 	<ul style="list-style-type: none"> • Crowding out effecten
Voor- en nameting	Goed	<ul style="list-style-type: none"> • Individueel niveau • Bedrijfsniveau • Sectorniveau • Regionaal niveau 	<ul style="list-style-type: none"> • Eventuele andere ontwikkelingen of interventies kunnen effectmeting vertroebelen

Uitkomsten waarop een evaluatie van de effecten van een leerrekening zich kan richten

Wat betreft de leerrekening kan gedacht worden aan de volgende uitkomstmaatstaven:

1. Bereidheid en geplande intensiteit om te gaan trainen
2. Stortingen en gebruik van budget
3. Feitelijke cursusdeelname
4. Werk- en leermotivatie
5. (Gepercipieerde) inzetbaarheid
6. Productiviteit
7. Beloning
8. Baanbehoud

3.2 Leerambassadeurs

Leerambassadeurs zijn werknemers die hun collega's inlichten over hun opleidingsmogelijkheden en hen motiveren hiervan gebruik te maken.¹⁸ Het is belangrijk dat leerambassadeurs collega's zijn die op hetzelfde functieniveau werken, en dus geen leidinggevenden zijn.¹⁹ In Engeland is er al veel ervaring opgedaan met leerambassadeurs. Sinds 1998 worden leerambassadeurs in Engeland ingezet om het leren onder werkenden te bevorderen. De vakbonden spelen hier een centrale rol en worden daarin ondersteund door de overheid. Ook in Nederland zien de vakbonden goede perspectieven voor het opzetten van leerambassadeurs. Zo hebben FNV en CNV met andere partners leerambassadeurs geïntroduceerd bij Philips.²⁰

¹⁸ Zie ook Coenen (2018).

¹⁹ Ook uit het onderzoek van Künn-Nelen e.a. (2018) bleek dat cursusdeelname onder laagopgeleiden positief gerelateerd is aan feedback van collega's. Dit gold niet voor feedback van leidinggevenden.

²⁰ Zie voor een volledige lijst Woudstra (2017).

Effectmetingen van leerambassadeurs

De effectiviteit van leerambassadeurs kan het beste gemeten worden met een trapsgevijs gerandomiseerd experiment maar ook een DD-methode en een voor-en nameting behoren tot de mogelijkheden.

Een RD-design leent zich daarentegen niet voor het evalueren van leerambassadeurs. Immers, iedereen heeft in principe toegang tot een leerambassadeur.

Een voor- en een nameting lijkt een eenvoudige manier om het effect van het introduceren van leerambassadeurs te evalueren. De scholingsdeelname voor en na de introductie van leerambassadeurs kan immers vergeleken worden. Echter, hiermee is nog niet aangetoond dat een eventuele toename van de cursusdeelname het gevolg is geweest van de introductie van de leerambassadeurs. Immers, als een bedrijf besluit leerambassadeurs in te zetten, dan zal het bedrijf vaak ook op andere manieren meer aandacht zijn gaan besteden aan LLO. Zo zou het kunnen zijn dat leidinggevenden meer aandacht zijn gaan besteden aan LLO bij functioneringsgesprekken. Het simpelweg vergelijken van cursusdeelname voor en na de introductie van leerambassadeurs kan dus een te positief beeld opleveren. Tegelijkertijd zou een dergelijke evaluatiemethode ook tot een onderschatting van het daadwerkelijke effect kunnen leiden als werknemers cursussen buiten het bedrijf om gaan volgen. Deze cursussen zijn waarschijnlijk niet in de bedrijfsadministratie opgenomen en komen dus alleen boven water wanneer een werknemerssurvey onderdeel uitmaakt van de evaluatie. Dit moet dan zowel voor als na de introductie gebeuren. Het bovengenoemde gevaar op overschatting blijft echter bestaan.

Dit gevaar is weg te filteren door een DD-methode toe te passen. DD-analyses controleren immers voor veranderingen die door andere factoren dan de interventie worden veroorzaakt. Er zijn een aantal mogelijkheden om met behulp van een DD-analyse de introductie van leerambassadeurs te evalueren. Zo kunnen binnen een bedrijf verschillende afdelingen vergeleken worden als niet in alle afdelingen/vestigingen een leerambassadeur wordt ingezet. Daarnaast kan ook op sectorniveau bekeken worden wat de effecten van de introductie van leerambassadeurs zijn. Bedrijven met en zonder een leerambassadeur zouden binnen een sector zowel voor als na de introductie vergeleken kunnen worden. De focus op verschillende afdelingen/vestigingen binnen een bedrijf of verschillende bedrijven binnen een sector vergroot de kans dat aan de common trend assumptie wordt voldaan. Ook kunnen sectoren waarbinnen leerambassadeurs wel ingezet worden, vergeleken kunnen worden met sectoren waar geen leerambassadeurs ingezet worden. Door minimaal twee voor- en twee nametingen te hebben kan het effect geïdentificeerd worden. Soms kunnen de voormetingen overigens ook op basis van historische data plaatsvinden en hoeft er met de invoering van leerambassadeurs dus niet gewacht te worden voordat de voormetingen afgerond zijn. Idealiter worden bedrijfsregistraties gebruikt, hoewel dit alleen mogelijk is bij objectieve uitkomsten zoals cursusdeelname.²¹ Indien men geïnteresseerd is in subjectieve en/of gepercipi-

²¹ Dit heeft wederom wel als beperking dat cursussen/opleidingen die gevolgd zijn buiten het bedrijf om niet meegenomen worden in de effectmeting.

eerde uitkomsten zoals de bereidheid van werknemers om aan scholing deel te nemen of de bewustwording van de noodzaak tot verdere ontplooiing kunnen de voormetingen niet retrospectief plaatsvinden.²²

Het voordeel van een trapsgewijs gerandomiseerd veldexperiment is dat voormetingen niet nodig zijn. Zo zou een bedrijf met meerdere vestigingen ervoor kunnen kiezen om op gerandomiseerde wijze leerambassadeurs op verschillende momenten in de tijd in zetten op de verschillende vestigingen. Een bedrijf kan dit onderbouwen doordat de financiële middelen en praktische mogelijkheden om bij alle vestigingen tegelijk te starten beperkt zijn. Op het moment dat bij de eerste vestiging de leerambassadeurs aan de slag gaan, krijgen de werknemers bij alle andere vestigingen ook een vragenlijst met daarin onder andere vragen over de bereidheid van werknemers om aan scholing deel te nemen en hun gepercipieerde noodzaak tot verdere ontplooiing. Een half jaar later starten andere vestigingen ook met leerambassadeurs en wordt een vergelijkbare vragenlijst afgenomen onder de werknemers uit alle vestigingen. Weer een half jaar later start de laatste groep vestigingen met leerambassadeurs en krijgt wederom iedereen een vragenlijst. In plaats van vestigingen kunnen overigens ook afdelingen binnen een bedrijf gekozen worden²³ of bedrijven binnen een sector. Cruciaal blijft wel dat de toewijzing van de vestigingen/afdelingen/bedrijven random plaatsvindt. Het feit dat wel alle vestigingen/afdelingen/bedrijven uiteindelijk leerambassadeurs kunnen introduceren zorgt ervoor dat een trapsgewijs gerandomiseerd veldexperiment hier de voorkeur heeft boven een gerandomiseerd veldexperiment.

Tabel 3.2 vat de mogelijkheden voor het evalueren van leerambassadeurs samen.

²² Tenzij deze informatie uit personeelsadministratie te halen is (bijv. uit verslagen van functioneringsgesprekken).

²³ Er mag alleen geen sprake zijn van spill-over effecten, waardoor dit alleen mogelijk is als de afdelingen niet nauw samenwerken.

TABEL 3.2 Overzicht gebruik evaluatiekader voor evaluatie leerambassadeurs

	Toepasbaarheid	Niveaus	Mogelijke obstakels
Grootschalig gerandomiseerd experiment	Beperkt	<ul style="list-style-type: none"> • Bedrijfsniveau (tussen vestigingen of afdelingen) • Regionaal niveau 	<ul style="list-style-type: none"> • Mogelijke ethische kwesties
Trapsgewijs gerandomiseerd veldexperiment	Zeer goed	<ul style="list-style-type: none"> • Bedrijfsniveau (tussen vestigingen of afdelingen) • Sectorsniveau • Regionaal niveau 	<ul style="list-style-type: none"> • Alleen mogelijk met gebruik van paneldata
DD-methode	Goed	<ul style="list-style-type: none"> • Bedrijfsniveau (tussen vestigingen of afdelingen) • Sectorsniveau • Regionaal niveau 	<ul style="list-style-type: none"> • Common trend assumptie nodig • Bedrijfsadministratie niet alomvattend wat betreft cursusdeelname • Minimaal twee voormetingen nodig
RD-design met voormeting	Nee		
RD-design zonder voormeting	Nee		
Voor- en nameting	Beperkt	<ul style="list-style-type: none"> • Bedrijfsniveau 	<ul style="list-style-type: none"> • Eventuele andere ontwikkelingen/interventies vertroebelen effectmeting • Bedrijfsadministratie niet alomvattend wat betreft cursusdeelname • Voormeting nodig

Uitkomsten waarop een evaluatie van de effecten van leerambassadeurs zich kan richten

Wat betreft de leerambassadeurs kan gedacht worden aan de volgende uitkomstmaten:

1. Gepercipieerde noodzaak tot leren
2. Aantal gesprekken met de leerambassadeur
3. Bereidheid en geplande intensiteit om te gaan trainen
4. Feitelijke cursusdeelname
5. Werk- en leermotivatie
6. (Gepercipieerde) inzetbaarheid
7. Productiviteit
8. Beloning
9. Baanbehoud

4 MONITORING TREND LLO IN NEDERLAND

Naast de evaluatie van specifieke LLO-interventies, is het ook belangrijk om inzicht te krijgen in de trend in het LLO in Nederland. Want waar het voor de effectmeting van één specifieke LLO-interventie van belang is dat de toename in bijvoorbeeld cursusdeelname alleen te wijten is aan die specifieke LLO-interventie, willen we ook weten of er door de focus op LLO in de brede zin ook een daadwerkelijke stijging is in de deelname aan LLO-activiteiten. Een trendanalyse is hiervoor geschikt. Hiertoe is er data nodig die over een langere periode in de tijd het LLO in Nederland meet. Hoewel het hier in eerste instantie gaat om cursusdeelname en informeel leren op het werk, zijn meerdere facetten belangrijk. Wat cursusdeelname betreft is het bijvoorbeeld interessant om te analyseren of over de tijd de financiering (en het initiatief voor het volgen) van de cursussen verschuift van bijvoorbeeld de werkgever naar de werknemer of naar O&O fondsen. Ook is het interessant om de leermotivatie van de Nederlandse potentiële beroepsbevolking en het leerklimaat binnen bedrijven over de tijd te volgen. Dergelijke trends kunnen in kaart gebracht worden door langlopende vragenlijsten te analyseren en voort te zetten. Een aantal voorbeelden van dergelijke vragenlijsten zijn:

4.1 Huishoudpanels en werknemersvragenlijsten

Allereerst zijn er verschillende relevante huishoudpanels en werknemersvragenlijsten.

LISS panel

Het Longitudinal Internet Studies for the Social sciences (LISS) panel is een representatief huishoudpanel.²⁴ Het bevat 4500 huishoudens met in totaal 7000 individuen. Panelleden krijgen elke maand een online vragenlijst die zij tegen betaling kunnen invullen. Er zijn een aantal jaarlijks terugkerende vragenlijsten en daarnaast is er binnen het LISS panel de mogelijkheid om andere data te verzamelen. Één van de jaarlijkse vragenlijsten die binnen het LISS panel wordt afgenomen is de Work & Schooling vragenlijst. Deze is inmiddels al 11 keer afgenomen en bevat vragen over cursus/opleidingsdeelname in de afgelopen 12 maanden, het type cursus/opleiding, cursus/opleidingslocatie, cursus/opleidingsduur, reden voor cursus/opleidingsdeelname en de financiering van de

²⁴ Voor meer informatie: <https://www.lissdata.nl/about-panel>

cursus/opleiding. Het LISS panel is daarmee geschikt om de trend van (aspecten van) cursus/scholingsdeelname in kaart te brengen.

ROA LLL Enquête

De ROA Levenslang leren Enquête (ROA LLL Enquête) monitort sinds 2004 de cursus-deelname, het informeel leren en de kennisontwikkeling in Nederland.²⁵ Elke drie jaar wordt de vragenlijst opnieuw afgenomen onder een representatieve steekproef van de Nederlandse bevolking.²⁶ In 2017 is de ROA LLL Enquête afgenomen binnen het LISS panel. Hierdoor is het mogelijk om op individueel niveau koppelingen te maken met andere informatie die binnen het LISS panel is verzameld. De ROA LLL Enquête bevat gedetailleerde vragen over cursus/opleidingsdeelname in de afgelopen 12 maanden: onderwerp cursus, cursusduur, cursus gevolgd in vrije of werktijd, zelfstudie onderdeel van de cursus, (erkende) certificering van de cursus, financiering van de cursus, initiatief voor cursusdeelname, belang van cursus voor verschillende situaties en redenen voor deelnemen cursus. Naast cursusdeelname wordt ook gevraagd naar informeel leren op het werk, zelfstudie en deelname aan conferenties of workshops.²⁷ Voor deze leervormen is gevraagd wat het belang was voor verschillende situaties. Ook komt de leermotivatie en het leervermogen aan bod.²⁸ Daarnaast zijn er vragen opgenomen over de tevredenheid van de loopbaanperspectieven, scholingsmogelijkheden en ontwikkelingsmogelijkheden binnen de huidige baan.²⁹ Er zijn ook een aantal vragen opgenomen die inzicht geven in de werkgeverskant van het leren. Zo is er een blok vragen over HR-instrumenten opgenomen om inzicht te krijgen in de manier waarop de werkgever aandacht besteedt aan het functioneren en/of de loopbaanontwikkeling van de werknemers. Daarnaast is het leerklimaat onderdeel van de ROA LLL Enquête.³⁰ De ROA LLL Enquête is door de specifieke focus bij uitstek geschikt om de trend van (aspecten van) cursus/scholingsdeelname, maar ook van informeel leren op het werk en van zelfstudie in kaart te brengen. Daarnaast kan op basis van de ROA LLL Enquête ook de trend in de gepercipieerde ondersteuning door de werkgever in kaart gebracht worden. Bij een continuering van de ROA LLL Enquête kan bovendien de trend in de leermotivatie en het leerklimaat geanalyseerd worden.

NEA

De Nationale Enquête Arbeidsomstandigheden (NEA) is vragenlijst onder een representatieve steekproef onder werknemers.³¹ Het betreft een cross-sectioneel onderzoek dat jaarlijks door 23.000 tot 45.000 werknemers wordt ingevuld. De vragenlijst bevat een vraag over (de reden voor) cursusdeelname in de afgelopen twee jaar en de huidige gepercipieerde behoefte tot het volgen van een cursus. Daarnaast wordt er aan respondenten gevraagd om aan te geven hoe belangrijk zij de mogelijkheid tot leren vinden en

²⁵ In Fouarge e.a. (2018) worden enkele LLO-trends inzichtelijk gemaakt.

²⁶ Voor meer informatie: <http://roa.sbe.maastrichtuniversity.nl/?portfolio=roa-lifelong-learning-survey>

²⁷ Informatie over deelname aan conferenties of workshops maakt sinds 2017 deel uit van de vragenlijst.

²⁸ Niet in alle metingen (op eenzelfde manier) opgenomen.

²⁹ Sinds 2017 in vragenlijst opgenomen.

³⁰ Sinds 2017 in vragenlijst opgenomen

³¹ Voor meer informatie: <https://www.monitorarbeid.tno.nl/nea>

hoe tevreden zij zijn over deze mogelijkheid in hun huidige baan. Ook wordt gevraagd of de leidinggevende de ontwikkeling van kennis en vaardigheden steunt. De NAE leent zich door zijn omvang met name voor het in kaart brengen van de trend in de cursusdeelname van verschillende groepen werknemers. Een nadeel ten opzichte van de eerdergenoemde datasets is dat de NEA het niet mogelijk maakt om mensen te volgen over de tijd waardoor het niet mogelijk is om aanvullende informatie te vergaren door middel van additionele vragenlijsten onder dezelfde respondenten.

EBB

De Enquête Beroepsbevolking (EBB) is een zogenaamd roterend panelonderzoek onder ruim 190.000 respondenten die vijf keer benaderd worden.³² In de EBB zijn een aantal vragen opgenomen over cursusdeelname in de afgelopen vier maanden, waarbij een onderscheid gemaakt wordt tussen cursussen die korter dan 6 maanden duren en cursussen/opleidingen die minimaal 6 maanden duren. Het betreft de inhoud van de cursus/opleiding, de cursus/opleidingsduur, reden voor cursus/opleidingsdeelname en de financiering van de cursus/opleiding, cursus gevolgd in vrije of werktijd en of er sprake was van betaald studieverlof. De EBB kan gebruikt worden om trends in de ontwikkeling van de deelname aan kortdurende versus langlopende cursussen in kaart te brengen.

³² Voor meer informatie: <https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/enquete-beroepsbevolking--ebb-->

Vignette onderzoeken voor input van beleid

Veel beleidsinstrumenten op het gebied van LLO moeten nog geïmplementeerd worden. Leerambassadeurs en leerrekeningen komen nog maar in beperkte mate voor. Bovendien is het de vraag hoe leerrekeningen precies vorm moeten krijgen en hoe we het gebruik kunnen optimaliseren. De vraag is bijvoorbeeld of werknemers op het moment dat zij een leerrekening tot hun beschikking krijgen ook überhaupt hier interesse voor hebben. En voor welke cursussen en opleidingen hebben zij eigenlijk interesse? En maken zij de juiste keuze?

Voordat een specifieke interventie wordt geïntroduceerd kan er eerst verkennend onderzoek worden gedaan door middel van surveys. Daarmee kan inzicht worden verkregen in de meest relevante doelgroepen (groepen mensen of bedrijfssectoren met een lage trainingsparticipatie) en de factoren die de deelname mogelijk kunnen bevorderen of belemmeren. Een zeer waardevol instrument dat een brug bouwt tussen gewone surveys en experimentele interventiemetingen is het zogenaamde *stated-choice* of *vignette onderzoek*. Vignette onderzoek kan al voor een groot deel antwoorden geven op de vragen die hierboven kort zijn aangestipt. Vignette onderzoek is een vorm van experimenteel onderzoek waarbij met gebruik van denkbeeldige situaties (vignettes) kan worden vastgesteld welke omstandigheden invloed hebben op de houdingen en gedragingen en wat de mogelijke gevolgen zijn van de invoering van een leerrekening of van bepaalde communicatietypen omtrent LLO. Een vignette onderzoek gaat veel dieper in op het verklaren van situaties dan een normale vragenlijst en maakt het mogelijk om de gevolgen van nog niet bestaande interventies te analyseren.

Een vignette onderzoek is waardevol vanwege diverse redenen:

1. Een vignette onderzoek kan in een meer simpele beleidscontext helpen om de werkelijke voorkeuren van mensen te achterhalen zelfs voordat de interventie in de praktijk wordt geïntroduceerd.
2. Een vignette onderzoek kan deze voorkeuren voor een groot deel meten zonder dat experimenten in de praktijk hoeven te worden ingevoerd.
3. Vignettes kunnen zo ontwikkeld worden dat mensen keuzes moeten maken die zij normaliter ook in de praktijk zouden maken. De uitkomsten van vignettenonderzoeken komen derhalve ook beter overeen met de resultaten van experimenten in de praktijk dan surveyresultaten.
4. Vanwege de exogene variatie in vignettes kunnen de effecten die geschat worden in een vignettenonderzoek ook causaal worden geïnterpreteerd.
5. Vignette onderzoeken zijn relatief goedkoop ten opzichte van echte experimenten.

Een nadeel is uiteraard dat vignette onderzoek uiteraard alleen gedrag in hypothetische situaties meet. Verschillende onderzoeken hebben systematisch onderzocht in hoeverre het hypothetische gedrag in vignetonderzoeken overeenkomt met feitelijk gedrag en vinden overtuigend bewijs dat dit het geval is (Peabody e.a., 2000, 2004; Eifler, 2007; Telser en Zweifel, 2007; Eggers, 2015).

Dergelijk vignettenonderzoek is reeds en kan ook in de toekomst worden geïmplementeerd in bestaande panels/vragenlijsten zoals de ROA LLL Enquête. Zie bijvoorbeeld Künn-Nelen e.a. (2018).

4.2 Werkgeversvragenlijsten

Daarnaast is het mogelijk om via langlopende bedrijfsenquêtes inzicht te krijgen in de trend van cursusdeelname en opleidingsuitgaven per bedrijf. De volgende datasets lenen zich hiervoor:

Onderzoek Bedrijfsopleidingen

Het onderzoek bedrijfsopleidingen is een 5-jaarlijks CBS onderzoek onder bedrijven met minimaal 10 werknemers in de particuliere sector.³³ Het onderzoek bedrijfsopleidingen geeft inzicht in de opleidingsactiviteiten van bedrijven. De belangrijkste onderwerpen zijn de scholingsparticipatie en de intensiteit van en uitgaven aan cursussen. Daarnaast worden bedrijven bevraagd over het opleidingsbeleid en de kwaliteitsborging van gegeven cursussen. Met behulp van het onderzoek bedrijfsopleidingen is het mogelijk om zicht te krijgen op de manier waarop de scholingsuitgaven van bedrijven zich over de tijd ontwikkeld.

ROA Werkgeversenquête Scholingsbeleid

In 2017 heeft het ROA een werkgeversenquête scholingsbeleid gehouden onder ongeveer duizend Nederlandse ondernemingen met minimaal vijf medewerkers. Respondenten werd gevraagd of zij de scholingsdeelname van verschillende groepen werknemers binnen het bedrijf te laag, te hoog of op het juiste niveau vonden. Daarnaast werd gevraagd wat de belangrijkste redenen zijn voor het bedrijf om in scholing en ontwikkeling van medewerkers te investeren. Ook werd het leer-klimaat van het bedrijf bevraagd, en werd expliciet naar de gepercipieerde leermotivatatie en het gepercipieerde leervermogen van verschillende groepen werknemers gevraagd. In de enquête zijn bovendien verschillende vignettenexperimenten opgenomen om inzichtelijk te krijgen in hoeverre werkgevers bereid zijn specifieke groepen werknemers scholing aan te bieden en onder welke omstandigheden. Omdat de ROA werkgeversenquête scholingsbeleid een eenmalige vragenlijst betreft is het zonder vervolgmeting niet mogelijk trends in kaart te brengen. Bij eventuele vervolgmetingen zou het daarentegen mogelijk zijn om te analyseren of er een ontwikkeling gaande is in de bereidheid van werkgevers om in verschillende groepen werknemers te investeren en in de door de werkgever gepercipieerde leermotivatatie en gepercipieerde leervermogen van hun werknemers.

Deze werknemers- en werkgeversvragenlijsten bieden mogelijkheden om evaluaties van LLO-interventies dus in de context te plaatsen van bredere trends in de LLO-ontwikkeling in Nederland. Een bijkomend voordeel van enkele van deze vragenlijsten is dat ze daarnaast ook de mogelijkheid bieden om vignettenexperimenten uit te voeren die een indicatie kunnen geven van de effecten van nog niet uitgevoerde LLO-interventies. Voor het verkrijgen van een zo goed mogelijk representatief beeld van de LLO-ontwikkeling en uitwerking van LLO-interventies verdient het de voorkeur om een combinatie van vragenlijsten te gebruiken.

³³ Voor meer informatie: <https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/bedrijfsopleidingen>

5 BEOORDELING EVALUATIEMETHODE VAN TOEKOMSTIGE PROJECTEN

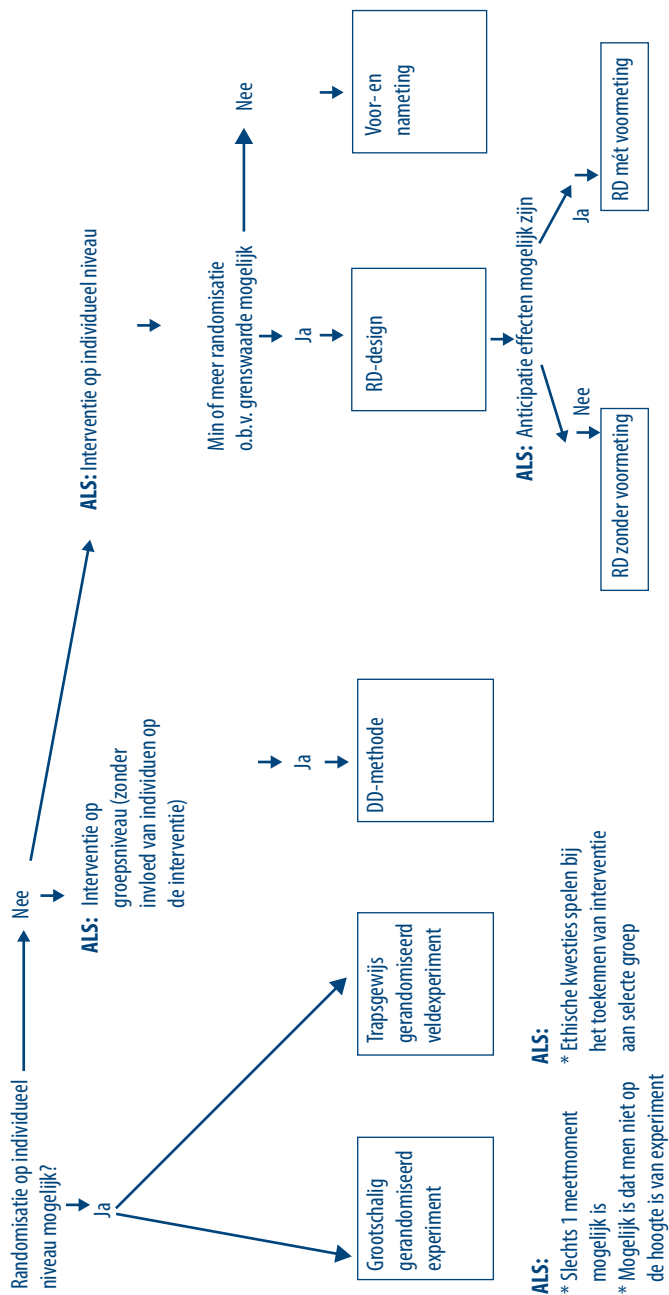
In dit afsluitende hoofdstuk presenteren we een praktische toolkit voor de beoordeling van de kwaliteit van de voorgestelde effectmeting in een projectvoorstel voor een LLO-interventie. Allereerst presenteren we een beslisboom (zie Figuur 1) die gebruikt kan worden door de indieners van een project. Deze beslisboom vergemakkelijkt de keuze voor een bepaalde methode, afhankelijk van de voorgestelde LLO-interventie. In aanvulling daarop wordt een overzicht gegeven van de voorwaarden waaraan moet worden voldaan als voor een bepaalde methode op het effect van de interventie te meten wordt gekozen (zie Tabel 5.1).

Daarna wordt een checklist gepresenteerd die gebruikt kan worden voor de goedkeuring van projecten. In deze checklist is een aantal vragen opgenomen die kunnen helpen bij het beslissen of projectvoorstellen al dan niet voldoen aan de eisen voor het goed uitvoeren van effectmetingen van LLO-interventies. Mochten één of meerdere vragen niet beantwoord worden, dan is een aanpassing van de opzet van het project en/of de effectmeting noodzakelijk.³⁵ Indien dit niet mogelijk is dan betekent dit dat interventiemetingen niet op betrouwbare wijze kunnen worden uitgevoerd. In dat geval zou een vignettestudie een goed alternatief kunnen zijn voor het daadwerkelijk introduceren van een LLO-interventie.

5.1 Beslisboom en voorwaarden waaraan voldoen moet worden als voor een bepaalde evaluatiemethode wordt gekozen

³⁵ Een asterisk geeft aan dat de evaluatie opgeschaald kan worden naar schaal 5 van de Maryland scale als de bijbehorende vraag met ja beantwoord kan worden.

FIGUUR 5.1 Beslisboom



TABEL 5.1 Voorwaarden waaraan voldaan moet worden als voor een bepaalde evaluatiemethode wordt gekozen

Grootschalig gerandomiseerd experiment	Trapsgewijs gerandomiseerd veldexperiment	DD-methode	RD-design	Voor- en nameting
1. Interventie mag niet gerelateerd zijn aan de uitkomstmaat	1. Interventie mag niet gerelateerd zijn aan de uitkomstmaat	1. Interventie mag niet gerelateerd zijn aan de uitkomstmaat	1. Interventie mag niet gerelateerd zijn aan de uitkomstmaat	1. Interventie mag niet gerelateerd zijn aan de uitkomstmaat
2. Selectieve respons moet kunnen worden uitgesloten	2. Selectieve respons moet kunnen worden uitgesloten	2. Selectieve respons moet kunnen worden uitgesloten	2. Selectieve respons moet kunnen worden uitgesloten	2. Selectieve respons moet kunnen worden uitgesloten
3. Men mag niet weten dat men onderdeel uitmaakt van een onderzoek/experiment	3. Men mag niet weten dat men onderdeel uitmaakt van een onderzoek/experiment	3. Men mag niet weten dat men onderdeel uitmaakt van een onderzoek/experiment	3. Men mag niet weten dat men onderdeel uitmaakt van een onderzoek/experiment	3. Men mag niet weten dat men onderdeel uitmaakt van een onderzoek/experiment
4. Spill-over effecten mogen niet mogelijk zijn	4. Spill-over effecten mogen niet mogelijk zijn	4. Spill-over effecten mogen niet mogelijk zijn	4. Spill-over effecten mogen niet mogelijk zijn	
5. Eventuele andere interventies moeten niet meer of minder betrekking hebben op de interventie- en controlegroep	5. Eventuele andere interventies moeten niet meer of minder betrekking hebben op de interventie- en controlegroep	5. Eventuele andere interventies moeten niet meer of minder betrekking hebben op de interventie- en controlegroep	5. Eventuele andere interventies moeten niet meer of minder betrekking hebben op de interventie- en controlegroep	
6. De toewijzing van individuen aan interventie- en controlegroepen moet willekeurig zijn.	6. De toewijzing van individuen aan interventie- en controlegroepen moet willekeurig zijn.		6. Grote N nodig	
	7. Grote N nodig			
	8. Langere looptijd nodig/ meerdere meetmomenten nodig			
		6. Minimaal 2 voor- en 2 nametingen vereist		
		7. Common trend vereist		

Grootschalig gerandomiseerd experiment	Trapsgewijs gerandomiseerd veldexperiment	DD-methode	RD-design	Voor- en nameting
			7. Variabele aanwezig waarop interventie- en controlegroep kan worden onderscheiden * op basis van discontinuïteit * moet continue variabele zijn * mensen mogen niet in staat zijn deze variabele te manipuleren * controleren voor trends noodzakelijk * optimale bandbreedte moet vastgesteld kunnen worden	
			8. Interventie en controlegroepen moeten op controlevariabelen identiek zijn	
				4. Geen trend in uitkomst-variabelen toegestaan
				5. Er mag maar één interventie tegelijk plaatsvinden
				6. Er mogen geen andere relevante ontwikkelingen plaatsvinden
				7. Minimaal 1 voormeting nodig (voor men op de hoogte is van de interventie)

5.2 Checklist voor goedkeuring projecten

1. Wordt één van de hier genoemde methoden toegepast?
2. Staat de geplande interventie los van de uitkomstmaat?
3. Indien survey data gebruikt gaat worden, kan selectieve respons uitgesloten worden?
4. Kan gegarandeerd worden dat deelnemende partijen géén relatie kunnen leggen tussen het geplande onderzoek en de interventie?
5. In geval van een grootschalig gerandomiseerd experiment:
 - a. Leidt de randomisatie tot een volledig willekeurige indeling in interventie- en controlegroepen?
 - b. Kan gegarandeerd worden dat eventuele andere interventies evenredig/willekeurig vertegenwoordigd zijn in de interventie- en de controlegroep?
 - c. Kan gegarandeerd worden dat er geen spill-over effecten kunnen plaatsvinden tussen de interventie- en controlegroepen?
6. In geval van een wijzing van een tragsgewijs gerandomiseerd veldexperiment:
 - a. Leidt de randomisatie tot een volledig willekeurige indeling in interventie- en controlegroepen?
 - b. Kunnen de interventie- en controlegroepen over de tijd gevolgd worden?
 - c. Is de steekproef groot genoeg om het verwachte effect te kunnen meten?
 - d. Kan gegarandeerd worden dat eventuele andere interventies evenredig/willekeurig vertegenwoordigd zijn in de interventie- en de controlegroep?
 - e. Kan gegarandeerd worden dat er geen spill-over effecten kunnen plaatsvinden tussen de interventie- en controlegroepen?
7. In geval van een DD-methode:
 - a. Wordt aan de common-trend assumptie voldaan? Hiervoor zijn minimaal 4 meetmomenten – 2 voor en 2 na de interventie – nodig.
 - b. Kan gegarandeerd worden dat eventuele andere interventies evenredig/willekeurig vertegenwoordigd zijn in de interventie- en de controlegroep?
 - c. Kan gegarandeerd worden dat er geen spill-over effecten kunnen plaatsvinden tussen de interventie- en controlegroepen?
8. In geval van een RD-design:
 - a. Is de toedeling van mensen aan de interventie- en de controlegroep zo goed als steekproefsgewijs verlopen en gebaseerd op een discontinuïteit?
 - b. Zijn de interventie- en de controlegroep identiek op de controlevariabelen?
 - c. Is de variabele op basis waarvan het onderscheid wordt gemaakt tussen de interventie- en de controlegroep een continue variabele?
 - d. Kan gegarandeerd worden dat eventuele andere interventies evenredig/willekeurig vertegenwoordigd zijn in de interventie- en de controlegroep?
 - e. Is er een optimale bandbreedte bepaald rondom de grenswaarde op basis waarvan de interventiegroep van de controlegroep wordt onderscheiden en is de steekproef groot genoeg binnen dit interval?
 - f. Kan er gecontroleerd worden voor trends in de variabele op basis waarvan de interventiegroep van de controlegroep wordt onderscheiden uitgesloten?

- g. Indien crowding out effecten mogelijk zijn, is een voormeting mogelijk?
 - h. Kan gegarandeerd worden dat eventuele andere interventies willekeurig vertegenwoordigd zijn in de interventie- en controlegroep?
 - i. Kan gegarandeerd worden dat er geen spill-over effecten kunnen plaatsvinden tussen de interventie- en controlegroepen?
9. In geval van een voor- en een nameting:
- a. Kan uitgesloten worden dat er meerdere interventies tegelijk plaatsvinden?
 - b. Kan de voormeting plaatsvinden voordat de deelnemende partijen op de hoogte zijn van de geplande interventie?
 - c. Kan gegarandeerd worden dat er tussen de voor- en nameting geen andere (voor de uitkomstvariabele) relevante ontwikkelingen plaatsvinden?
 - d. Kan er aannemelijk gemaakt worden (idealiter door data) dat er geen trend is in de uitkomstvariabele?

Literatuur

Coenen (2018). Vakbondswerk maken van scholing en ontwikkeling. *Zeggenschap* 2018/3, pp. 40-42.

De Grip, Lindeboom & Montizaan (2012). Shattered Dreams: The Effects of Changing the Pension System Late in the Game. *The Economic Journal*, Vol. 122 (559), pp. 1-25.

De Grip & Sauermann (2012). The effects of training on own and co-worker productivity: evidence from a field experiment. *The Economic Journal*, Vol. 122, pp. 376-399.

Eggers, Fowler, Hainmueller, Hall & Snyder (2015). On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races. *American Journal of Political Science*, Vol. 59(1), pp. 259-274.

Eifler (2007). Evaluating the validity of self-reported deviant behavior using vignette analyses. *Quality & Quantity*, Vol. 41(2), pp. 303-318.

Farrington, Gottfredson, Sherman & Welsh (2002). Maryland Scientific Methods Scale. In: Sherman et al. (Eds.), *Evidence-based crime prevention* (p. 13-21). London: Routledge

Fouarge, Van Eldert, De Grip, Künn-Nelen, Poulissen (2018). *Nederland in leerstand*. ROA-R-2018/4, Maastricht.

Gerards, De Grip & Witlox (2014). "Employability-miles" and worker employability awareness. *Applied Economics*, Vol. 46, pp. 952-965.

Görlitz (2010). The effect of subsidizing continuous training investments—evidence from German establishment data. *Labour Economics*, Vol. 17(5), pp. 789-798.

Hidalgo, Oosterbeek & Webbing (2014). The impact of training vouchers on low-skilled workers. *Labour Economics*, Vol. 31, pp. 117-128.

Künn-Nelen, Poulissen, Van Eldert, Fouarge & De Grip (2018). *Leren onder werkenden met een kwetsbare positie op de arbeidsmarkt*. ROA-R-2018/5, Maastricht.

Leuven & Oosterbeek (2004). Evaluating the effect of tax deductions on training. *Journal of Labor Economics*, Vol. 22(2), pp. 461-488.

Montizaan & Fouarge (2016). The impact of employer's characteristics and beliefs on the propensity to hire older job applicants: Evidence from a stated preferences experiment. *Netspar DP 01/2016-054*.

Peabody, Luck, Glassman, Dresselhaus & Lee (2000). Comparison of vignettes, standardized patients, and chart abstraction: A prospective validation study of 3 methods for measuring quality. *Journal of the American Medical Association*, Vol. 283(13), pp. 1715-722.

Peabody, Tozija, Muñoz, Nordyke & Luck (2004). Using vignettes to compare the quality of clinical care variation in economically divergent countries. *Health Services Research*, Vol. 39, pp. 1951-970.

Schwerdt, Messer, Woessmann & Wolter (2012). The impact of an adult education voucher program: Evidence from a randomized field experiment. *Journal of Public Economics*, Vol. 96, pp. 569-583.

Telser en Zweifel (2007). Validity of discrete-choice experiments evidence for health risk reduction. *Applied Economics*, Vol. 39(1), pp. 69-78.

Woudstra (2017). *Union Learning Reps in Nederland*. Verleden, heden en toekomst.